Assessment and Improvement of Unfamiliar Face Processing and Identity Verification in Security and Forensics

Thesis (cumulative Thesis)
presented to the Faculty of Arts and Social Sciences
of the University of Zurich
for the degree of Doctor of Philosophy

by Mirjam Fuhrer

Accepted in the fall semester 2016
on the recommendation of the doctoral committee:
Prof. Dr. Klaus Oberauer (main advisor)
Prof. Dr. Janek Lobmaier

Zurich, 2016

Abstract

Face processing is an important ability in several occupations in the security domain. Recognition and identity verification of individuals play a crucial role for instance in criminal investigations or passport control. Whereas humans are very accurate in recognizing familiar faces, processing unfamiliar faces is highly error-prone.

In this thesis several approaches to improve accuracy in these tasks are discussed. Two empirical studies are presented. The first study introduces a newly developed test, which is specifically tailored to the requirements needed in security related occupations. There are remarkable individual differences in the ability of perceiving or recognizing unfamiliar faces, which can hardly be affected by training or practice. The test measures face perception as well as face memory and can serve as a method in the selection of security personnel. The study contains a comparison of the performance of security personnel and laypeople.

The second study aimed at an improvement of identity verification by introducing the shape of ears as a new biometric feature in addition to the face. As results revealed, human observers are able to match individual faces presented on frontal photographs if only the ears are available as visual information. Accuracy in simultaneous matching of unfamiliar faces was improved in two experiments when attention to the ears was increased. A simple instruction to pay attention to the ears could be helpful in applied settings such as passport control.

Zusammenfassung

Die Verarbeitung von Gesichtern ist in vielen Berufen im Sicherheitsbereich eine wichtige Fähigkeit. Die Erkennung von Personen und die Überprüfung ihrer Identität spielt beispielsweise in strafrechtlichen Ermittlungen oder bei der Passkontrolle eine wichtige Rolle. Während Menschen sehr gut in der Erkennung von bekannten Gesichtern sind, ist die Verarbeitung von unbekannten Gesichtern höchst fehleranfällig.

In dieser Arbeit werden verschiedene Ansätze zur Verbesserung der Genauigkeit in diesen Aufgaben diskutiert. Die Arbeit beinhaltet zwei empirische Studien. Die erste Studie stellt einen neu entwickelten Test vor, der speziell an die Anforderungen an Berufe im Sicherheitsbereich angepasst ist. In der Fähigkeit, unbekannte Gesichter zu erkennen und unterscheiden gibt es grosse Unterschiede zwischen verschiedenen Personen, die mit Training oder Übung kaum beeinflusst werden können. Der Test misst Gesichtswahrnehmungs- und Gesichtsgedächtnis und kann für die Personalselektion von Sicherheitspersonal eingesetzt werden. Zusätzlich beinhaltet die Studie einen Vergleich der Leistung von Sicherheitspersonal und Laien in verschiedenen Gesichtserkennungsaufgaben.

In der zweiten Studie wurde die Strategie, die Ohrenform als biometrisches Merkmal zusätzlich zum Gesicht zu nutzen, untersucht. Die Ergebisse zeigten, dass Beobachter fähig sind, Personen auf Porträtfotografien nur anhand ihrer Ohren zu unterscheiden. Die Genauigkeit im simultanem Bildvergleich konnte in zwei Experimenten verbessert werden, wenn die Aufmerksamkeit auf die Ohren erhöht wurde. Eine einfache Instruktion, die Ohren zu beachten, könnte hilfreich sein für angewandte Bereiche wie beispielsweise Passkontrollen.

Contents

1 Synopsis

1.1 Background

Over the last two decades, security has become ubiquitous in political debate, media, as well as in research. Recent terroristic activities have led to higher interest in national and international security but also to concerns about how well-equipped law enforcement agencies are to respond to the changing dynamics of security in contemporary society. (Mythen & Walklate, 2016)

As a reaction to the new threat situation, large investments have been made into modern security technology as well as in approaches to train human operators in this technology (Schwaninger, 2005; Schwaninger, Hofer, & Wetter, 2007). In airport security, which has increased in prominence over the past few years, passengers are increasingly screened by means of several technologies such as X-Ray devices to screen hand luggage, liquid explosive detection systems or body scanners (Hofer & Wetter, 2012; Wetter & Fuhrer, 2013). Moreover, several states have invested in so-called behaviour detection programs aiming at recognizing suspicious behaviour (Koller, Wetter, & Hofer, 2015a). The majority of these efforts to improve security and to prevent crimes and terroristic activities are concentrated on attempts to pick unknown and anonymous individuals out from a crowd, based on their suspicious behavior or the possession of prohibited items.

Despite these technological processes, or new behavioral detection programs, the identification or verification of identity is still of great importance in the security domain. Related issues are homeland security, surveillance, access control or border protection. This has led to an increased interest in biometrics. Biometrics can be defined as unique measurable physiological or behavioral characteristic that can be used to confirm or determine the identity of an individual. There are many possible biometrics such as DNA, fingerprints, hand geometry, retinal scans, gait, facial structure and all of them have its advantages and disadvantages in application. (Jain, Ross, & Prabhakar, 2004)

The International Civil Aviation Organisation (ICAO) has decided that the facial picture remains the primary biometric trait for biometric passports. The ICAO reached this conclusion based on the several advantages applied to facial images compared to other biometrics: The use of facial images is already socially and culturally accepted internationally and the public is already aware of the capture of facial images for the use of identity verification purposes. In addition, human verification of a person against a photograph is relatively simple and a familiar process for border control authorities. The capture of a facial image is non-intrusive and the end user does not have to touch or interact with a physical device. Furthermore, facial image capture can be deployed both immediately but also retrospectively and many states have a legacy database of facial images. A facial image can be captured from an endorsed photograph, not requiring the person to be physically present. For watch lists, a photograph of the face is generally the only biometric available. (ICAO, 2015)

Computer as well as human face processing is the most easy accessible way to identify a person, regardless if the person is aware of the identification attempt and whether she is cooperative or not. However, human as well as technological ability to perform these tasks accurately is limited. The next paragraph introduces concrete examples of tasks in the security and forensic domain, which are based on face processing.

1.1.1 Tasks in security involving face processing

Faces are used as means of providing identity in several forensic and security applications. The tasks which rely on face processing can be separated broadly into the two domains identification and identity verification. In verification or authentication, a system or a human verifies, if a person is who she claims to be, based for instance on identity documents. Identification is a more complex task, in which the identity of an unknown individual is of interest. (Introna & Nissenbaum, 2010)

Accuracy in these tasks is not only important to hinder crimes and catch criminals, but also to prevent innocent people of the consequences of wrongful identifications based on errors in identity verification.

Concrete tasks related to identification are performed in criminal investigations when missed or wanted persons are spotted by police officers on patrol. Furthermore, eyewitnesses of a crime sometimes provide descriptions of the criminal, and the face is the most individual and distinctive feature which is visible most of the time. It is known, that eyewitness testimony in lineups or construction of faces are not very reliable, due to several reasons (for a review see Lindsay, Mansour, Bertrand, Kalmet, & Melsom, 2011). Further, surveillance of certain areas is more and more common. A few years ago, there were approximately 4 million closed-circuit television (CCTV) devices in the UK (Security Newsdesk, 2013) and 30 million in the USA (Davis & Valentine, 2009). In the German speaking society, surveillance by the means of CCTV is less common, as there are more concerns about privacy and data protection. However, airports and other important public areas are monitored by cameras as well (Koller, Wetter, & Hofer, 2015b). Sometimes, a crime is recorded with CCTV. This evidence can be useful in two situations. A person displayed in CCTV footage can be released to the public and recognized by someone familiar to them, referring to identification. Second, CCTV evidence can be used at court to match the identity of a defendant, referring to identity verification. (Bobak, Hancock, & Bate, 2016).

There is a possibility to consult an expert, if the suspect denies to be depicted on the photo. In this case, facial image specialists such as forensic anthropologists evaluate the similarity of the pictorial evidence to the suspect using various techniques. They then report their opinion based on several technologies, such as anthropometry, in which distances between facial features are measured. Besides criminal investigations, there are several occasions in which identity verification is accomplished, typically by matching a live present person with a picture in an identity document. The most important example is passport control along borders. Another common application is entrance control of buildings. Moreover, identity documents with photos are presented in many common settings, such as checking personal tickets in public transportation, in stores to verify ages when buying alcohol or students presenting their university cards to get discounts. In all of these situations, personnel perform the simultaneous matching of a person’s face with the picture in their document.

Considering the described tasks, it is obvious, that face processing plays an important role for law enforcement agencies, such as police forces. To improve the accuracy in tasks relevant for security, it is possible to either improve technological equipment or human performance (Schwaninger, 2005; Wetter, 2012). Although the empirical studies in this thesis focus on human face processing, the current state of face recognition technology is shortly discussed in the next section.

1.1.2 The limits of face recognition technology

There is a huge industry looking at the automation of face recognition for security application (Bruce, 2011; Introna & Nissenbaum, 2010). An example of the developments of automated face recognition systems would be e-Gates. An increasing number of airports installed automated border control technologies, also called e-Gates. Most of these systems rely on face recognition. Typically, the system captures a facial photograph of the traveler and conducts a biometric verification against the image stored on the biometric passport chip. The algorithms of these systems are usually kept secret by the vendors as their intellectual property. The gate opens, if the recognition is successful, otherwise the passengers is directed to manual passport control. (Labati et al., 2015)

The e-gates are expected to improve security, convenience for passengers and efficiency of the border crossing and reduce labor costs. However, the human operator cannot be replaced for the following reasons; currently, in European countries the use of e-gates is restricted to citizens of states taking part in the Schengen cooperation who hold biometric identity documents (Labati et al., 2015). Additionally, some eligible passengers are unable or reluctant to use the e-gates (Oostveen, 2014; Oostveen, Kaufmann, Krempel, & Grasemann, 2014). Furthermore, in order to guarantee recognition accuracy of the e-gates, the capture of a high quality picture is of major importance. However, many aspects, such as illumination differences due to environmental changes, or difficulties to require the user to stand correctly in front of the camera, cause problems (Labati et al., 2015; Spreeuwers, Hendrikse, & Gerritsen, 2012). Another issue is that there have been some cases of so-called presentation attacks to face recognition systems, in which a person presents a picture of the legitimate enrollee or, more sophisticated, wears a printed face mask (Raghavendra, Raja, & Busch, 2015). Currently, automated face recognition systems are not able to recognize if a mask or a picture is presented instead of a real face (Hengfoss, Mull, Püschel, & Jopp van Well, 2015).

In the task of matching identity in pairs of full frontal face images, the best algorithms were in range of human performance already almost e decade ago (O'Toole et al., 2007). However, in these tests the pictures did not vary in viewpoint and the backgrounds of the pictures were standardized (O'Toole, 2011). Moreover, the task of identity verification is computationally easiest to achieve for face recognition systems as only two facial pictures have to be compared (Introna & Nissenbaum, 2010). At present, automatic face recognition technology is highly accurate in constrained environments but human are still superior in complex person identification tasks with variable conditions such as the recognition of a person with unstandardized backgrounds, different viewing angles or from video footage (Best-Rowden, Bisht, Klontz, & Jain, 2014; Phillips, Hill, Swindle, & O'Toole, 2015) Similarly, human still outperform technology in face image retrieval, in which the task is to find all images containing the same person from a dataset of faces (Zhou & Li, 2015).

Even finding a face in an image which is effortless for human is computationally complex and not trivial to achieve (Hancock, Bruce, & Burton, 2000). Therefore, application of face recognition systems for person identification of suspects or unknown individual is very challenging. A recent example of an unsuccessful attempt to employ face recognition technology is the Boston Marathon Bombings in 2013. Despite extensive video footage from local surveillance systems and onlookers cameras, law enforcement agencies were unable to identify the two suspects with face recognition technology (Klontz & Jain, 2013). According to Introna & Nissenbaum, 2010, a surveillance scenario in which a face is picked out of a crowd in an uncontrolled environment is unlikely to become an operational reality for face recognition technology in the foreseeable future.

In summary, despite considerable advances in face recognition algorithms, machines cannot yet replace humans in several tasks related to face processing at the moment. In addition, even when face recognition technology is employed, human operators almost always have the final say in identification judgements (White, Dunn, Schmid, & Kemp, 2015). As Schwaninger, 2005 states, technological equipment is of limited value if the humans who operate it are not selected and trained to perform their task accurately and efficiently. Thus, training and selection of personnel in the security domain concerning face processing is the main topic of this thesis. The aim of the studies presented in this thesis was to develop a tool to select personnel according to their face processing abilities and to search for methods to improve face processing. Human face processing is therefore discussed in the next section.

1.2 Human face processing

There was a long debate whether there are unique visual mechanisms for processing the identity of faces compared to other objects. It is now widely accepted, that faces are special compared to other objects (for a review see McKone & Robbins, 2011).

A number of cognitive models of face identity processing have been proposed (Breen, Caine, & Coltheart, 2000; Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Valentine, 1991). A discussion of these models would be beyond the scope of this thesis but according to Wilhelm et al., 2010 there are two processes assumed to be involved in most of them. The first process is the perception of faces, the second refers to face memory. When a face is perceived, *pictorial codes* are derived from the retinal input, these are relatively raw images. Then, the *structural codes*, which are descriptions of the face independent from viewpoint or expression, are extracted. These are required to distinguish individual faces and they contain information of facial features and their arrangement. First-order features include facial elements such as nose size and shape. Second-order or configural features include the spacial relationsships of the first-order features such as the distance between nose and eyes. Some authors have proposed that faces are processed holistically and the features of a face and their spatial relations are fused together in an unified representation (Tanaka & Farah, 1993; Tanaka & Gordon, 2011). However, Schwaninger, Lobmaier, & Collishaw, 2002 could show that featural and configural information both play an important role for face recognition and that these two sources of information can also be encoded and stored independently of one another.

The second process is face memory, which is the ability to remember facial stimuli and depends on face recognition units (FRUs). They can be seen as interconnected sets of structural codes for each face which are stored within long-term memory (Bruce & Young, 1986). During the visual process of face perception, the structural codes are compared to the ones stored in long-term memory. A face is recognized as familiar, when both structural codes match. Recognizing a face thus requires the maintenance of structural codes stored within face recognition units, the comparison of stored and currently perceived facial structures and the correct retrieval of the corresponding face recognition unit. (Wilhelm et al., 2010).

Perception is therefore a prerequisite of memory. However, it is assumed in this thesis, that face memory and face perception are two separate abilities. Research (e.g., Herzmann, Danthiir, Schacht, Sommer, & Wilhelm, 2008) has shown, that performance in these two abilities is related to a certain degree. However, there are some people with exceptionally good performance in face memory, but not in face perception tasks (Bobak, Bennetts, Parris, Jansari, & Bate, 2016). As some results are only applicable to face memory or only to face perception, it is important to separate these two terms. Applied to the before mentioned face processing tasks which take place in security and forensic settings, identity verification commonly requires *face perception*, whereas tasks in which a person has to be identified mostly rely on *face memory*. One task which is difficult to categorize within this differentiation are simultaneous lineup tasks in which a target face is simultaneously presented with a lineup consisting of another picture of the target and several distractors (e.g., Megreya & Burton, 2008). This is an identification task which relies on face matching. Some terms need further clarification. The noun *face recognition* is often used as an umbrella term for face perception as well as face memory in psychology and for identity verification as well as identification when technology is considered. However, recognition sometimes refers to the activity concerning face memory in comparison to face perception as the word directly derived from memory (i.e., memorize) has a different meaning. *Face processing* or *face cognition* are other umbrella terms involving face memory and face perception. For clarification, in this thesis the term *face memory* is used when an identification task is meant. The meaning of the term *face perception* is in general psychological research broader than *face matching* and involves for instance processes of social cognition such as the assessment of attractiveness or attribution of certain personality traits based on the appearance of a face (Bruce & Young, 2012). However in this thesis, *face perception* means only the perception of identity, and the ability to discern faces presented simultaneously. Therefore the two terms *face matching* and *face perception* are used interchangeably.

1.2.1 Human performance in face memory and face perception

The ability to recognize a person from their facial appearance is essential for social interaction. Humans' ability to process, recognize and perceive faces familiar to them even from footage of very bad picture quality is impressive (Burton, Wilson, Cowan, & Bruce, 1999). However, the ability to process unfamiliar faces is far less accurate and much more influenced by variations such as differences in age (Megreya, Sandford, & Burton, 2013), expression (Bruce, 1982; Henderson, Greenwood, Hancock, Burton, & Miller, 1999) or viewpoint (Liu & Chaudhuri, 2002; Meinhardt-Injac, Meinhardt, & Schwaninger, 2009). This concerns not only the recognition of a stranger from memory but also simple perceptual tasks such as the simultaneous matching of a person to a photograph.

It has been assumed, that the fact that identity verification in security settings relies on human face matching despite the high error-rates which have been observed for unfamiliar faces in the field (Kemp, Towell, & Pike, 1997) as well as laboratory studies (e.g., Bruce et al., 1999; Burton, White, & McNeill, 2010; Henderson, Bruce, & Burton, 2001) is based on an overconfidence in our ability to process unfamiliar faces stemming from the impressive accuracy with which familiar faces are processed (Burton & Jenkins, 2011). In relation to the tasks performed in the security and forensic domain, unfamiliar faces are of particular importance. The findings discussed below thus refer to the processing of unfamiliar faces.

Face perception and memory abilities of security personnel such as police officers or passport control officers are important. There are in general two ways to improve performance within these professions. First, specially skilled people can be selected, for special positions in which face processing plays an important role either from the general public or from within the own ranks. Second, personnel can be trained. The choice of which process is preferable depends on several factors and questions such as: Are there large individual differences? Is the ability trainable and to what degree? The next sessions therefore discusses previous research considering individual differences, trainability and effects of expertise.

1.2.2 Individual differences, expertise and trainability

There are large individual differences in the ability to remember (Russell, Duchaine, & Nakayama, 2009) and to match faces (Megreya & Bindemann, 2013) ranging from nearly perfect to almost guessing probability.

On one end of this range lie developmental prosopagnosics, on the other end super-recognizers. Developmental prosopagnosia is a condition marked by exceptionally poor face processing with an absence of other cognitive deficits. Research estimates that about 2% of the general population has the condition (Duchaine & Nakayama, 2005; Kennerknecht et al., 2006). Super-recognizers in contrast are people with exceptionally good face processing. Russell et al. (2009) coined the term super-recognizers after they tested four individuals who were as good at face processing as developmental prosopagnosics are bad. Super-recognizers are not thoroughly researched yet, as the phenomenon has only recently been discovered. A recent study with a detailed cognitive assessment of the face- and object processing abilities of six super-recognizers revealed mixed results in relation to the question, if the skills of these people are restricted to face-specific processes. More specifically, super-recognizers also varied in whether face perception was also enhanced additional to face memory (Bobak, Bennetts, et al., 2016; Bobak, Dowsett, & Bate, 2016).

As already mentioned, most research supports the hypothesis that face processing is a highly specialised skill. This theoretical standpoint is supported by findings that some individuals with prosopagnosia only have difficulties with face memory tasks and no deficits with other cognitive domains (Busigny, Joubert, Felician, Ceccaldi, & Rossion, 2010; Duchaine & Nakayama, 2005; Rossion, 2014). Additionally, research conducted with the general population has failed to find a relationship between face memory and performance on tests of non-facial visual or verbal memory (Dennett et al., 2011; Hildebrandt, Wilhelm, Schmiedek, Herzmann, & Sommer, 2011; Wilmer et al., 2010, 2012) as well as with general cognitive abilities such as intelligence, reasoning or processing of other visual objects (Hildebrandt et al., 2011; Zhu et al., 2010). Furthermore, monozygotic twins are more similar than dizygotic twins in face perception (Zhu et al., 2010) and face memory (Wilmer et al., 2010; Zhu et al., 2010) performance, suggesting that genetic factors could play a significant role in face processing. However, as (Zhu et al., 2010) note, that their data also suggest substantial environmental influence.

Considering expertise, there have not been many studies comparing experts in face processing with the general public and the results of these few studies have been mixed. This could be partly due to the fact that all people have experience in processing faces as it is a fundamental task in daily life. However, some attempts have been made to compare experts, such as security professionals who perform unfamiliar face recognition and perception as part of their job with lay people. White, Kemp, Jenkins, Matheson, & Burton (2014) tested Australian passport officers and could show that they were not better in simultaneous face matching than students. However, in another study, a subgroup of passport control officers responsible for facial comparison performed better than other passport control officers (White, Dunn, et al., 2015). Burton et al. (1999) demonstrated that a group of UK police officers with experience in forensic identification were not superior in performance to untrained students. Studies with experts for forensic image comparison have shown that they can outperform students in face matching (Norell et al., 2015) and identification tasks (Wilkinson & Evans, 2009) also when images are presented only shortly (White, Phillips, Hahn, Hill, & O'Toole, 2015). However, their work with anthropometry has also be criticized for not being reliable (Kleinberg, Vanezis, & Burton, 2007).

Pure practice with a task seems not to improve performance. White, Kemp, Jenkins, Matheson et al. (2014) could not find a relationship between face matching accuracy and employment duration in their study with passport control officers. Although especially forensic image specialists seem to show better performance with face processing than the general public, the underlying mechanisms are unclear and motivational effects and self-selection are as likely as an explanation for these results as practice effects. As individual differences within groups are reported to be large and seem to be stable over time (White, Kemp, Jenkins, Matheson, et al., 2014), pre-occupation personal selection seems to be a tool to be considered.

To discuss training studies in unfamiliar face processing, three factors are important to organize the results. First, it has to be differentiated, whether healthy participants or people with certain deficits such as prosopagnosia or autism were trained. In this thesis, training of people with deficits is not further discussed as it is of minor importance for face processing in the security field.

Second, it has to be discerned, if the training aimed at improvements in face memory or in face perception. Whereas attempts to train face memory consistently were unsuccessful (e.g., Dolzycka, Herzmann, Sommer, & Wilhelm, 2014; Malpass, Lavigueur, & Weldon, 1973; Woodhead, Baddeley, & Simmonds, 1979), improvements in face perception seem possible to a certain degree. These training successes were based on trial-by trial feedback (White, Kemp, Jenkins, & Burton, 2014) or observed after participants performed a face perception task in groups (Dowsett & Burton, 2015). However, the improvements were small and did not exceed the large individual differences. Additionally, not all attempts to improve face perception are successful. One strategy based on a classification of the shape did not improve face matching accuracy. This is particularly critical as the strategy is a common component of training programs carried out by several government agencies to improve the ability to detect identity fraud. (Towler, White, & Kemp, 2014)

Third, some studies trained specific stimuli for specific participants, for example the recognition of Japanese faces in Caucasian participants (Goldstein & Chance, 1985). Human remember and perceive own-race faces more accurately than other-race faces (for a review of the own-race bias see Meissner & Brigham, 2001). Processing of other race faces seems to be dependent from experience with the faces of this race at least to a certain degree and some training procedures were successful in improving recognition of other race faces (e.g., Stahl, Wiese, & Schweinberger, 2008). Some researchers assume simple perceptual differences in face processing and report successful reduction of the own-race bias by simply shifting attention to other facial features which they assume to be more diagnostic in other race faces (Hills, Cooper, & Pake, 2013; Hills & Pake, 2013). However, in an own unpublished experiment, we were not able to replicate those findings. Clearly, there is more research needed to develop and evaluate training techniques for other race faces. Although training of other race faces would be interesting also for the security domain, a comprehensive discussion of the different theories explaining the own race bias or of training procedures which has been tested is beyond the scope of this thesis. Some training studies aimed at training the recognition of own-race as well as other race faces. Whereas the performance with the faces of another race could be improved, there was no effect of the training for own-race faces.

(Elliott, Wills, & Goldstein, 1973; Goldstein & Chance, 1985). These findings are in line with the mentioned unsuccessful attempts to train face memory in general.

1.2.3 Possible ways to improve performance

Given the scale of the individual differences in face memory and face perception ability and the limited impact of experience and training, the most obvious strategy to optimize performance and improve the accuracy of the system is personnel selection (Bobak, Hancock, et al., 2016).

This approach was chosen from the Metropolitan Police of London, which has set up a specialised *super-recogniser unit*. Members are recruited from within the police and perform several tasks in which face processing is especially important, such as the identification of live suspects and from cameras (Jaslow, 2013). Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016 recently examined members of this special unit and reported that they perform at levels above control groups in several face matching tasks. However, they tested only four people of the unit consisting of approximately 200 officers (Jaslow, 2013). It remains unclear why only such a small sample was tested. Moreover, the procedure which was used to select exactly these four participants remains unmentioned. It is not clear, thus, if the performance of this unit is above performance of laypeople in relation to face cognition.

As personnel selection is not a satisfactory solution for all settings or situations, training is another approach to improve face processing. However, based on the findings discussed above, it can be assumed that it is not possible to improve face memory ability in general. In contrast memory of other race faces can be improved by training and face perception seems to be trainable to a certain degree. These aspects could be taken into account for training procedures.

One approach would be to take advantage of the mentioned factors which are known to improve face perception. Training procedures could involve trial by trial feedback or working in groups (Dowsett & Burton, 2015; White, Kemp, Jenkins, & Burton, 2014) or concentrate on the processing of other race faces. As mentioned, more research is needed to develop training procedures and understand underlying mechanisms.

Another possible approach is to use information additional to the face to improve identity verification. This strategy circumvents the difficulties associated with face perception trainings and the disadvantages faces have as biometrics. Computer scientists for example suggest multimodal biometric recognition for biometric systems such as face recognition combined with iris scans (Faundez-Zanuy, 2005). As biometric passports contain digital fingerprints, one strategy could be to provide passport control corners with automated fingerprint recognition systems additional to face recognition. Most biometrics require specialized technological equipment as they are difficult or impossible to process for human. However, some biometrics could be trainable for human operators. One opportunity is offered by the visual appearance of the ears. The ear has been proposed as promising passive biometric because it has reliable and robust features which are also extractable from a distance (Burge & Burger, 1996). Multimodal technological recognition using both ear and face resulted in a significant improvement over either individual biometric (Chang, Bowyer, Sarkar, & Victor, 2003). As ears are visible in some passport pictures, the usage of ears could also improve human identity verification.

In the next section, the two empirical studies reported in this thesis are shortly summarized. Both studies aspire to the improvement of identification accuracy of security personnel.

1.3 Empirical studies

The first study reported in the thesis, aimed at measuring face cognition performance of security professionals. We introduce a newly developed test, the Zurich Facial Cognition Test (ZFCT) which can be applied in personnel selection of security professionals. In addition, we examined if there are differences in performance in face memory as well as face perception between laypeople and security professionals. We tested subsamples, namely laypeople, police officers and passport control officers. To date this is the most extensive comparison of performance between security personnel and laypeople using different face cognition tasks. The Zurich Facial Cognition Test (ZFCT) which was newly developed is tailored to the specific requirements of face processing tasks in the security field. The test uses stimuli involving factors which are commonly encountered in face processing tasks in security and forensic settings and makes them particularly difficult, such as changes in viewpoint, disguised people, age differences and faces of other ethnicities. The goal of this approach was to mirror the challenges of face cognition in the reality of security domain professions as closely as possible. The test measures face perception and face memory with three different tasks. Additionally, two well-known tests were used in the study; the short version of the Glasgow Face Matching Test (Burton et al., 2010) for face perception and the more difficult version of the Cambridge Face Memory Test (Russell et al., 2009). The results revealed that the difficulty level of the ZFCT is high enough to avoid ceiling effects and to enable the measurement of highly-skilled performance in the range of super-recognition. A confirmatory factor analysis of all five tasks further showed that the two latent factors face perception and face memory were highly correlated, but not the same. Further, we found group differences between the tested subsamples in tasks related to face perception. In tasks related to face memory, we found no differences between security personnel and laypeople.

The second study aimed at improvement of identity verification with a pioneering strategy. Some of the police officers in study one reported to pay attention to the shape of the ears to decide whether two pictures depict the same person or not. Although it is well known in forensic anthropology and biometric computer science that the shape of the ears are unique and stable over time and can serve as a reliable feature to identify people, psychological research has paid no attention to the potential of ears for identification purposes so far. The approach of this study was to use another biometric feature in addition to the face. As training procedures in face processing are time-consuming and costly, as well as not very successful, we circumvented these challenges in testing a strategy which is independent from individual differences in face perception.

In experiment 1 of this study, we investigated if in a population of laypeople human observers are even able to distinguish people reliably based on the shape of their ears. We presented intact faces, faces without ears and faces in which only the ears were visible to compare the ear and the face as different sources of information. The results revealed that humans are indeed able to perform simultaneous matching of people only based on the shape of their ears. Moreover, accuracy was on a comparable level with stimuli, in which only the face without the ears was available as visual information. Additionally, performance with intact faces was higher after an instruction to pay attention to the ears than before, but it remained unclear, whether the instruction alone was effective or experience with the trials in which only ears were visible further boosted performance or trust in the strategy.

In the second experiment we therefore only used intact faces. To monitor initial attention to the ears and later compliance with the instruction, eyetracking was deployed in this experiment. We tested laypeople in two blocks following directly after each other. The experimental group received an instruction to pay attention to the ears between the two blocks, whereas the control group received no instruction. The results concerning eyetracking revealed that participants in general did not pay much attention to the ears initially. Participants of the experimental group increased the attention to the ears after the instruction without neglecting the face. Additionally, the instruction was effective in improving performance of the experimental group.

2 Measuring face matching and face memory performance of security professionals

Mirjam Fuhrer[1] and Franziska Hofer[2]

[1] University of Zurich, Cognitive Psychology Unit

[2] Kantonspolizei Zürich (Zurich State Police), Research & Development

Submission status

Submitted for publication.

Authors' contributions

MF: Review of the literature, design of the study, development of the test, selection of the stimuli, selection and preparation of the stimuli, data collection, data analysis and interpretation, writing of the manuscript

FH: Supervision and discussion of MF's contributions, revising the manuscript

Acknowledgment

This research was funded by a grant from the Swiss Federal Office of Civil Aviation and the Kantonspolizei Zürich (Zurich State Police). Special thanks go to the police officers and passport control officers from the Kantonspolizei Zürich who provided us with insight into their daily work and experience. The authors also thank Amanda Planzer for her assistance and Klaus Oberauer, Colin Wright and Haley Pruitt for revising the article. Thanks go also to the researchers who provide the images of their databases to the scientific community. Special thanks go to Jisien Yang and Sarah Chiller-Glaus, Corinne Koller and the foundation of Hans and Suzanne Biäsch for making us unpublished face databases available. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

Abstract

This study introduces the Zurich Facial Cognition Test (ZFCT) which has been constructed from stimuli with naturally occurring factors aggravating facial recognition without artificial changes of the photographs. The test measures face memory as well as face perception with three different tasks. The test is specifically tailored for the requirements of security related occupations The difficulty level is high enough to avoid ceiling effects and the test can also be used to measure highly-skilled performance in the range of super-recognition. Laypeople as well as police officers and passport control officers were tested with the ZFCT in addition with the short version of the Glasgow Face Matching Test (Burton et al., 2010) and the more difficult version of the Cambridge Face Memory Test (Russell et al., 2009). Police officers showed a better performance in the tasks closer related to face perception than face memory. Additionally, passport control officers showed a better performance in the simultaneous matching part of the ZFCT. The test is freely available for scientific use upon request.

Keywords:

face cognition, matching, face perception, face memory, security, personnel selection, security professionals, super-recognizer

2.1 Introduction

2.1.1 Face cognition

Face cognition is an essential socio-cognitive skill which is often taken for granted. Humans are described as face experts, because we can effortlessly recognize faces of many different people we know. Recognition of familiar faces is actually remarkably accurate even under conditions that make recognition difficult. In contrast, the recognition of unfamiliar faces is very error-prone. The difference between the recognition of familiar and unfamiliar faces was shown for example with low-quality videos from surveillance cameras (Bruce et al., 1999; Burton, Wilson, Cowan, & Bruce, 1999). Even a matching task, requiring a mere perceptual decision whether two simultaneously presented pictures depict the same person or not, can be difficult with unfamiliar faces (Henderson et al., 2001). These problems are not restricted to the comparison of photographs of people. The performance also remains error-prone with video instead of pictures (Bruce et al., 1999). Even having the person physically present does not necessarily make the task easier. In a field study, Kemp et al. (1997) showed that supermarket cashiers had difficulties judging whether pictures on credit cards depicted a simulated customer. Similarly, Davis & Valentine (2009) found high error rates in an eyewitness study including a condition where a live model was present in a room and had to be matched to a simultaneously presented video. Furthermore, it has been known for a long time that eyewitness identification is often highly unreliable (Borchard, 1932).

Herzmann et al. (2008) established a three factor model of face cognition involving face perception, face memory and speed of face processing. Face perception refers to the ability to perceive facial features and their configuration accurately and is best measured with perceptual tasks not requiring memory capacity. Face perception is typically measured by means of face matching tasks; participants have to decide, whether simultaneously presented photographs depict the same person or different people. Face memory is the ability to learn faces and retrieve them from memory for recognition. Measures of this ability involve memorization and a subsequent recognition of faces. A third factor is the speed of these processes, which is measured best with easy tasks. According to Herzmann et al. (2008) face perception and face memory are separable but highly related, whereas the speed of face cognition is only weakly correlated to the accuracy of face perception and face memory.

2.1.2 Individual differences in face cognition abilities

People vary to a great extent in their face cognition abilities (e.g., Woodhead & Baddeley, 1981). Prosopagnosia, a selective impairment in recognizing faces, represents the lowest tail of the distribution and has a prevalence of about 2% in the population (Kennerknecht et al., 2006). At the other end of the distribution are people with exceptionally good face cognition abilities, so-called super-recognizers (Russell et al., 2009).

Effects found in training studies with healthy subjects were small and of questionable durability for face perception (e.g., Dowsett & Burton, 2015; White, Kemp, Jenkins, & Burton, 2014). According to White, Kemp, Jenkins, Matheson, et al. (2014) passport officers are not better in simultaneous face matching tasks where they have to decide if two pictures show the same or two different people than laypeople despite their everyday experience with the task. Dolzycka et al. (2014) found that training had no effect on the accuracy of face memory. There is also evidence that individual differences in face memory are primarily due to genetic differences (Wilmer et al., 2010; Zhu et al., 2010).

2.1.3 Face cognition abilities in the security domain

Face recognition and face perception are important abilities for several tasks of security personnel (e.g. Chiller-Glaus, Schwaninger & Hofer, 2007). Police officers, for example, have to memorize faces from photographs when searching for missing people or people with arrest warrants. Often, the only reference is a CCTV image of very low quality. Police or passport control officers perform face perception whenever they have to verify an identity from an official document with a photograph. Therefore, individual differences in face cognition should be considered whenever it comes to hiring employees for positions in which face recognition or face perception is important.

2.1.4 Existing tests and aim of this test

As faces can change over time remarkably, e.g. through different hair styles and aging processes, unchanged photographs of faces should be used in order to boost ecological validity. This means no deliberate exclusion of facial features, such as hair or ears, the use of color instead of grey-scale photographs and no artificially generated faces. The use of natural stimuli should also be favored in future experiments to generalize results. According to Burton (2013), generalization also plays an important role for the future progress of face cognition research in general. To assess the two different mechanisms of face cognition, face perception and face memory, it would also be preferable to have a test which measures both mechanisms separately. Existing tests have not considered all of these criteria.

Two older tests, the Warrington Recognition Memory for Faces (Warrington, 1984) and the Benton Facial Recognition Test (Benton, Sivan, Hamsher, Varney, &

Spreen, 1994) have been criticized for including so much face-irrelevant information that no face cognition abilities are necessary to solve the trials (Duchaine & Weidenfeld, 2003). Tests with familiar faces like the Famous Faces Test (Fast, Fujiwara, & Markowitsch, 2004) neglect the differences in prior knowledge of the famous people. Moreover, studies suggest that familiar and unfamiliar faces are processed in different ways (for a review see Johnston & Edmonds, 2009) and security personnel deal most often with unfamiliar rather than familiar faces. Some existing face recognition tests have been developed to distinguish normal from subnormal face recognition or perception ability for the clinical field and are not difficult enough to differentiate between good and very good performers. In the Glasgow Face Matching Test (GFMT, Burton et al., 2010) mean performance is at approximately 90%. The Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006) is in large part constructed of faces without external features and involves pictorial noise to make the recognition more difficult. Super-recognizers also achieved perfect accuracy in the CFMT, and a more difficult version with even more noise has been developed to examine their performance (Russell et al., 2009). Herzmann et al. (2008) constructed a comprehensive test battery with several different tasks to measure all components of face cognition abilities but used only artificially generated faces.

The Zurich Facial Cognition Test comprehends photographs incorporating all these features that are present in face recognition in real life in general and in the security domain in particular. External features like hair and ears were not removed, with the aim of having test trials which are as realistic as possible. For the same reason, no artificially generated or changed face stimuli were used and all pictures were left colored and not converted to grey-scale. The aim of the test is to provide a tool which is tailored to the special requirements of security personnel. To investigate whether performance differences exist between security personnel and laypeople, three subsamples of participants were tested, namely: laypeople, police officers and passport control officers.

2.2 Method

2.2.1 Participants

A total of 195 participants (98 female, 97 male) with a mean age of 32.2 years (*SD* = 10.7) took part in this study. The sample consisted of three subsamples; laypeople, police officers and passport control officers. 104 laypeople (62 female) with a mean age of 28.1 (*SD*= 8.2) years were recruited via mailing list or online marketplace at the university. They volunteered to participate either for course credits or a small fee. Additionally, 65 police officers (19 female) with an average age of 31.3 years (*SD* = 7.7) who participated in a training course for the police took part in the study. The tests were part of a two day-training workshop. Furthermore, 26 passport control officers with a mean age of 50.4 years (*SD* = 6.3) volunteered to participate in the study for a small fee.

2.2.2 Materials and Procedure

The test includes three subtests consisting of different tasks (see the following section for a detailed description of the subtests). Each test part begins with an instruction page being presented on the screen. After the instruction the experimental trials follow immediately, without any practice trials.

In addition to the test described here, participants in our study solved the short form of the Glasgow Face Matching Test (Burton et al., 2010) previous to the ZFCT and the long form of the Cambridge Face Memory Test (Russell et al., 2009) after the ZFCT. The order of tests was the same for all participants.

Facial stimuli. In a first step, 555 face photographs were obtained from several different sources. The database used were: The FEI Database (De Oliveira Junior & Thomaz, 2006), the TarrLab face database (Righi, Peissig, & Tarr, 2012) the FERET database (Phillips, Moon, Rizvi, & Rauss, 2000; Phillips, Wechsler, Huang, & Rauss, 1998), the Minear & Park Face Database (Minear & Park, 2004), the Indian Face Database (Jain & Mukherjee, 2002), the VADANA face database (Somanath, Gowri & Kambhamettu, 2011), the FG-NET ageing database which is available upon request (for an overview of research using the database see Panis & Lanitis, 2014). In addition, some faces were retrieved from the public Interpol database of wanted people (Interpol, n.d.) and the FBI most wanted database (U.S. Federal Bureau of Investigation, n.d.). Additionally two unpublished face databases were used: The National Chung Cheng University Taiwanese Face Database from Jisien Jang and a Database from Corinne Koller and Sarah Chiller-Glaus, University of Zurich. No picture was used twice. External features like hair or ears were not removed from any of the pictures. Except for 19 pictures stemming from Swiss ID's, all photographs are in color and not converted to gray-scale. Yip and Sinha (2002) could show that color cues can play a role in face recognition under certain circumstances, and Bindemann and Burton (2009) showed that performance declines when color is removed. Backgrounds were only removed if they were overly distracting or when the two pictures from the same person could be identified because of the same background. In addition, clothing which made the target identifiable was removed and replaced with a white background using Photoshop CS4. Different realistic factors were varied in order to ensure a high difficulty level. According to Johnston & Edmonds' (2009) summary of the existing literature the following factors decrease the recognition of unfamiliar faces: Differences in view, expression, context or lightning, as well as negation and inversion. Negation and inversion are not naturally occurring and context is difficult to test, but different viewpoints, expression or lightning of the pictures or people are often significant in tasks which are carried out by security personnel and are therefore taken into account in the test. A further special challenge of real-life face cognition in the security domain is that identity documents can be up to ten years old, whereas face cognition tests usually use pictures of people recorded on the same day. Megreya et al. (2013) could show that face perception is much more difficult when pictures of people are months apart than when they are recorded on the same day. The ZFCT uses time differences in most simultaneous matching trials reflecting the real challenge of face perception in security settings as closely as possible. In some trials, a disguise is used resembling the attempt of criminals to hide facial features. Mansour et al. (2012) showed that identification accuracy decreased with the degree of disguise in a simultaneous and a sequential lineup task and therefore has an effect on face perception as well as face memory. Moreover, the test also includes faces of people from non-Caucasian ethnicities, which are more difficult to process for Caucasian participants (for a meta-analysis of the own-race bias see Meissner & Brigham, 2001) because security personnel have to deal with people of different ethnicities. For the learning phase of the facial memory test part a dark blue background is used. The faces are depicted either in frontal, three-quarter or profile view. In 11 trials photographs of people wearing a disguise or clothing like glasses, sunglasses, caps, scarfs or wigs are included.

Test development: Pilot studies. To obtain the final set of trials, a larger collection of trials were investigated in several pilot studies. Test items which could be solved by all participants and items which reached a mean correct score below guessing probability were removed. Sequential lineup and memory trials in which a certain distractor was more often chosen than the target itself were also excluded.

2.2.3 Final test set

Simultaneous matching. The final version of this test part consists of a total of 50 trials. Participants have to indicate whether two photographs depict the same person or two different people. The task is designed to minimize any influence of memory or recognition. Thus, the face stimuli to be compared are presented simultaneously on the screen. Both pictures are presented next to each other and are of the same size. Participants have 20 seconds to answer. Nineteen out of the 50 trials include an older passport picture paired with a more recent color photograph. Twenty trials are composed of Asian, 22 of Caucasian and 8 of African faces, 16 trials depict female and 34 male faces. Twenty-three trials are matches, showing two pictures of the same person, 27 are mismatches, showing photographs of two different people. For 18 matches there is an age difference of more than 2 years between the two pictures, for 4 matches, the person had changed hairstyle or was wearing glasses in one of the pictures. The trials are presented in random order. For an example of a match and a mismatch trial see Figure 1.

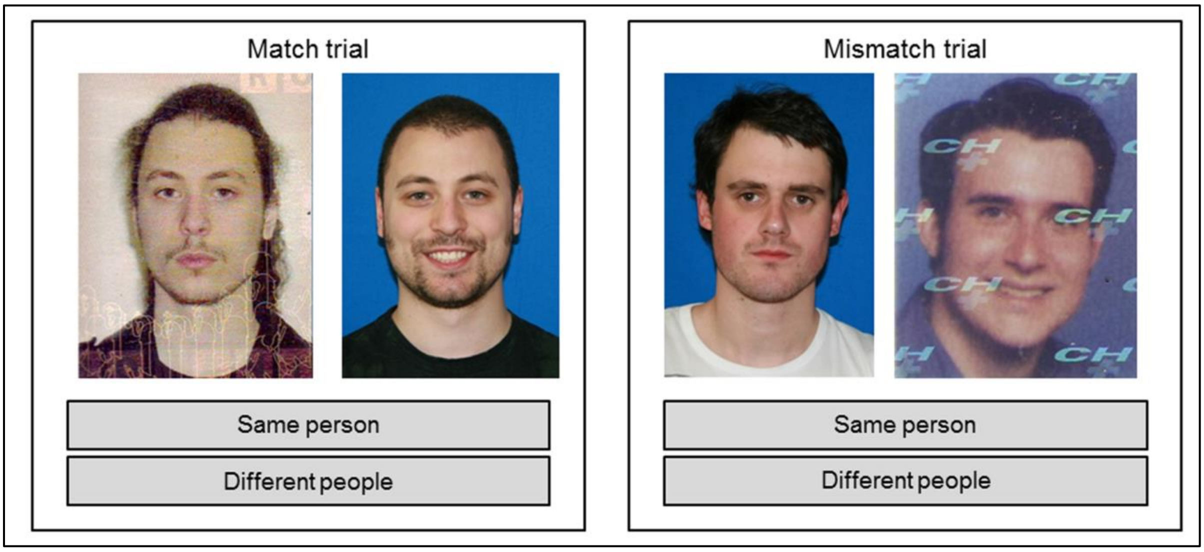

Figure 1. Examples of simultaneous matching trials

Sequential lineups. The task in this part is a sequential matching with a simultaneous lineup, hereinafter referred to as sequential lineup. A target face is presented for each set up to a maximum of five seconds, and participants are instructed to memorize the face. It is possible to shorten the target presentation by pressing a button. Next, a simultaneous lineup with five faces is presented. The target can be present or absent in the lineup. Approximately 1/6 of the trials are target absent trials (9 out of a total of 55 trials). Participants can either indicate the absence of the target by clicking on a “not present” button or click with the mouse on a photograph to indicate which one depicts the target person. The lineup is presented up to a maximum of 15 seconds. Participants can answer at any time. After 15 seconds, the faces disappear from the screen, but it is still possible to answer by clicking on the placeholder on the screen. For three trials, female faces are used, 40 trials are composed of Caucasian faces, 12 of Asian and 3 of African faces. In five trials, either the target or the lineup is created with disguised people, wearing sunglasses and caps. The trials are presented in random order. Likewise, the position of each face in the lineup is random. Figure 2 depicts an example of a sequential lineup trial.

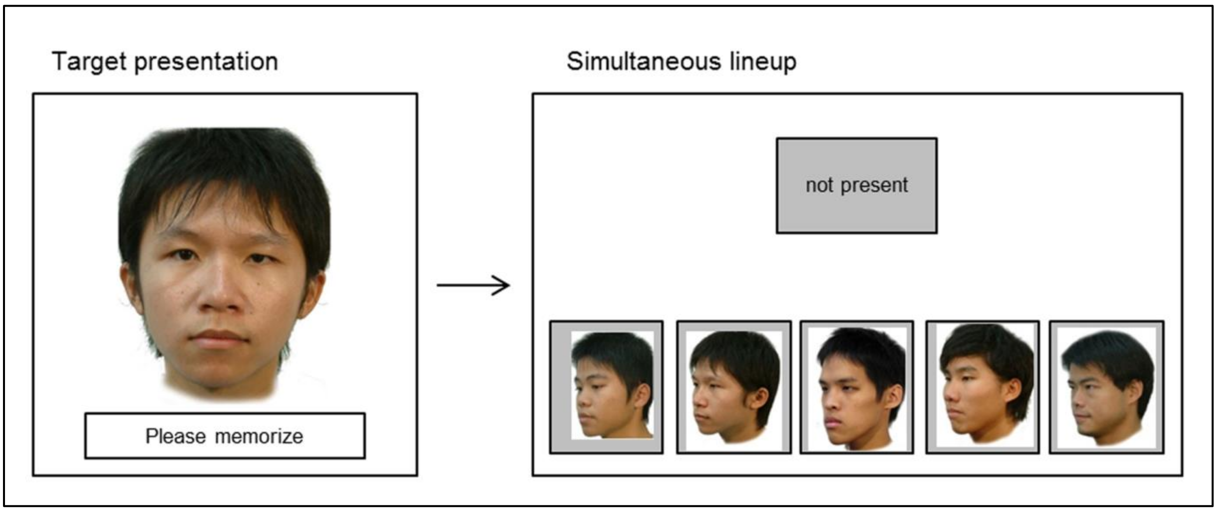

Figure 2. Example of a sequential lineup trial.

Face Memory. In the learning phase, eight male target faces are presented sequentially in frontal view. Five of the target faces are Caucasian and three Asian. Participants are given five seconds to view each image. Following this learning phase, participants are presented up to a maximum of 15 seconds with trials containing three faces. At any time, participants can indicate by mouse click either which face depicts the target face. If participants believe that the target face is not present, they can click on a "not present" button. Five out of the 39 trials are target absent trials depicting three new faces which were not shown in the learning phase. The trials are separated in an easier first part and a more difficult second part. The first part consists of 16 trials with pictures in frontal view followed by pictures in profile or three-quarter view. After these 16 trials, a second learning phase takes place in which the target faces are presented again in the same way as in the first learning phase. Then in the second part 23 new trials are presented. To maintain the difficulty despite repeated presentation of the targets, photos from identity cards, with disguise, significant age difference and low quality pictures taken from a movie similar to CCTV are used in the second part.

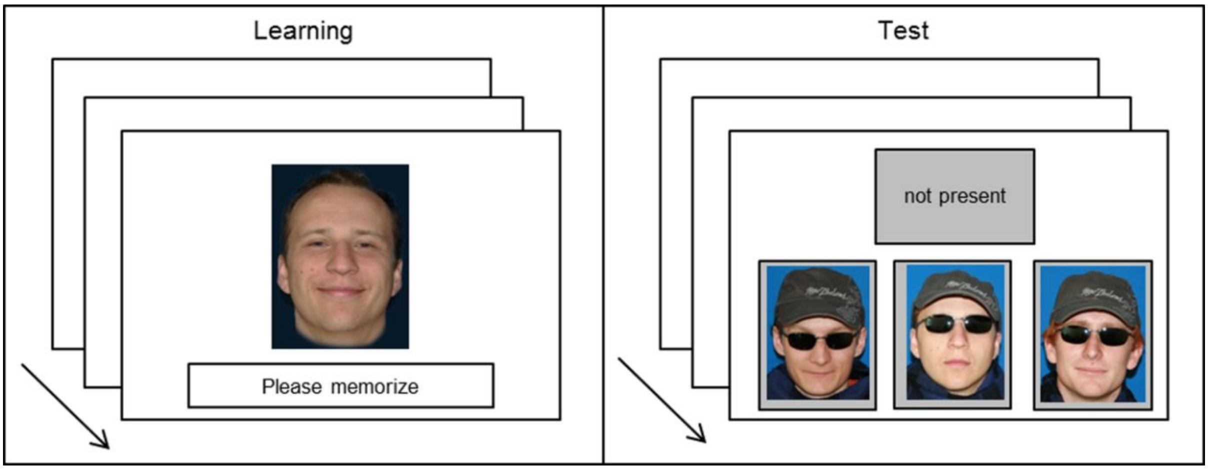

Figure 3. Schematic of the ZFCT memory task. In the learning phase, 8 targets were presented followed by several test trials.

2.3 Results

2.3.1 Data analysis

P-values and Bayes Factors. The widespread and traditional use of p-values in research has been criticized and there are several problems associated with null-hypothesis testing (e.g., Wagenmakers, 2007). Recently the American Statistical Association published a statement addressing the issue that p-values are still current practice despite several shortcomings (Wasserstein & Lazar, 2016). To take this ongoing debate into account, additional to the p-values, Bayes factors are reported in this paper. The Bayes factor is a ratio that compares the likelihood of the data under two different hypotheses, typically the null and the alternative hypothesis (for a short introduction in to Bayes statistics see e.g. Jarosz & Wiley, 2014). Shortly, an inverse Bayes factor (BF_{10}) reflects the ratio of the probability favoring the alternative hypothesis over the null hypothesis. For example a BF_{10} of 20 means that the data are twenty times more likely under the alternative hypothesis than the null hypothesis. A BF_{10} below 1 favors the null hypothesis H_0 and is reported as BF_{01}, reflecting the ratio of the probability of the null hypothesis compared to the alternative hypothesis. To interpret the Bayes factors, a verbal description is also provided based on the table of (Jeffreys, 2000). We calculated all bayesian analyses using JASP (JASP Team, 2016) and the default priors were used. Classical analyses were calculated with R (R Core Team, 2014).

2.3.2 Performance

GFMT short. Accuracy ranged from 47.5% to 100% correct answers, with a mean of 80.1% (SD = 10.1%). The guessing probability of this task is 50%. There was no correlation between accuracy and age of participant (r= -.013, p = .86; BF_{01} = 11). There was also no effect of gender [male 79.6%, female 80.6%; t(193) = .62, p = .48; BF_{01} = 5].

Zurich Facial Cognition Test: Simultaneous matching. Accuracy ranged from 42.0% correct to 94.0% correct, with a mean of 72.3% (*SD* = 9.9%). The guessing probability of this task is 50%. There was no correlation between accuracy and age of participant (r= .10, p = .16; BF_{01} = 4.2) and there was no effect of gender [male 72.8%, female 71.9%; $t(193)$ = .62, p = .53; BF_{01} = 5.3].

Zurich Facial Cognition Test: Sequential lineups. Accuracy ranged from 25.5% correct to 89.1% correct, with a mean of 61.0% (*SD* = 11.6%). The guessing probability of this task is 16.6%. There was no correlation between accuracy and age of viewer (r= .-.06, p = .37; BF_{01} = 7.5) and there was no effect of gender [male 61.0%, female 60.9%; $t(193)$ = .06, p = .95; BF_{01} = 6.4].

Zurich Facial Cognition Test: Face Memory. Accuracy ranged from 12.8% correct to 84.6% correct, with a mean of 46.3% (*SD* = 14.5%). The guessing probability of this task is 25%. There was a significant negative correlation between accuracy and age of participant (r= -.18, p < .01), older participants showed a tendency to perform worse than younger ones. The evidence for the alternative Hypothesis H_1 is only weak within the Bayesian framework (BF_{10}= 2.5). There was no effect of gender [male 44.6%, female 47.9%; $t(193)$ = -1.58, p = .12; BF_{01} = 2.0].

CFMT+. Accuracy ranged from 37.3% to 94.1% correct answers, with a mean of 63.3% (*SD* = 12.3%). The guessing probability is 33.3%. There was a significant negative correlation between accuracy and age of viewer (r= -.217, p > .01) with older participants exhibiting a tendency to perform worse than younger ones. The evidence for H_1 is substantial within the Bayesian framework (BF_{10} = 8.68). Additionally there was a significant difference between men and women [male 61.0%, female 65.5%; $t(193)$ = -2.6, p = .01]. The evidence is in substantial favor of the H_1 (BF_{10}= 3.58).

2.3.3 Relationships between face cognition tasks

There were significant and strong relationships between all face cognition tasks (Table 1). Thus, performance on the face perception tasks is related to performance in face memory tasks.

Table 1

Pearson correlations and Bayes Factors for accuracy rates among face cognition tasks

			ZFCT			
		GFMT short	**Simultaneous matching**	**Sequential lineups**	**Face Memory**	**CFMT+**
GFMT short	Pearson's r	—	.483 ***	.422 ***	.323 ***	.387 ***
	BF_{10}	—	9.612e +9 ***	1.263e +7 ***	3408.98 ***	464116.8 ***
Simultaneous matching	Pearson's r		—	.442 ***	.263 ***	.273 ***
	BF_{10}		—	9.713e +7 ***	85.60 **	146.7 ***
Sequential lineups	Pearson's r			—	.456 ***	.524 ***
	BF_{10}			—	4.358e +8 ***	1.344e +15 ***
Face Memory	Pearson's r				—	.507 ***
	BF_{10}				—	5.234e +13 ***
CFMT+	Pearson's r					—
	BF_{10}					—

Note. GFMT short = Glasgow Face Matching Test short Version; CFMT+ = Cambridge Face Memory Test long version; ZFCT = Zurich Facial Cognition Test.
* $p < .05$. ** $p < .01$. *** $p < .001$ for pearson correlations; * $BF_{10} > 10$. ** $BF_{10} > 30$. *** $BF_{10} > 100$ for Bayes factors. Bayes factors test the size of the correlation against zero.

The memory related tasks (ZFCT Face Memory and CFMT) showed stronger correlations among each other than to face perception tasks (GFMT and ZFCT simultaneous matching). The sequential matching part of the ZFCT showed a similar strong relationship to all memory and perception related test tasks. To examine the relations closer, a confirmatory factor analysis was conducted with lavaan (Rosseel, 2012). See figure 4 for the tested models and Table 2 for the model fit indices.

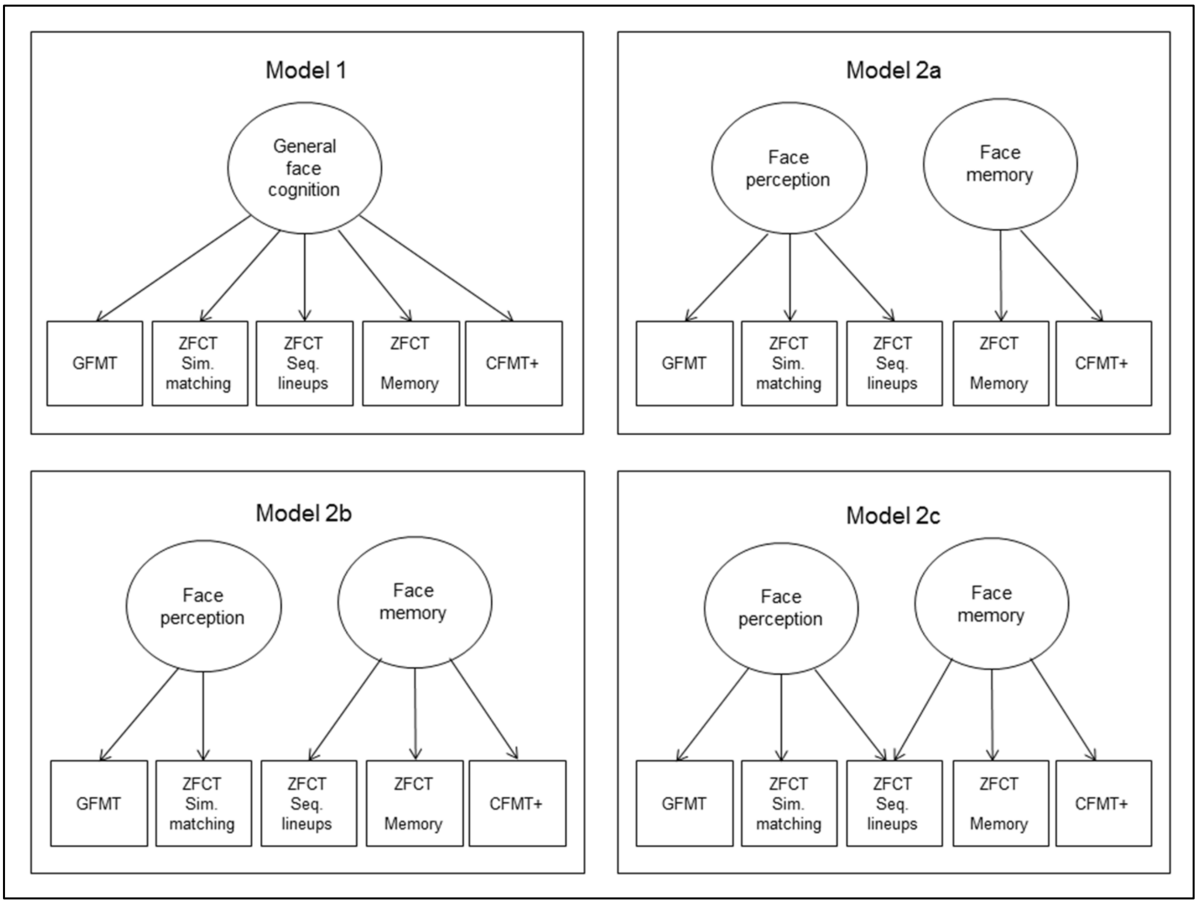

Figure 4. Models 1 to 2c tested in the confirmatory factor analysis. GFMT = Glasgow Face Matching Test short version; ZFCT = Zurich Facial Cognition Test; CFMT+ = Cambridge Face Memory Test long form.

Table 2
Goodness-of-Fit Indicators for Models 1 to 2c

Model	χ^2	df	χ^2/df	GFI	RMSEA	SRMR
1	24.3 ***	5	4.9	.95	.14	.06
2a	16.2 **	4	4.1	.97	.13	.05
2b	9.4	3	3.1	.98	.08	.03
2c	4.2	3	1.4	.99	.04	.02

Note. *** $p < .001$, ** $p < .01$, * $p < .05$

The data fitted better to the models with two latent variables face perception and face memory than to model 1 with only one general face cognition latent variable. From the models with two latent factors, Model 2c, in which the ZFCT sequential matching accounted for both latent variables face perception and face memory was best supported by the data $\chi^2(3) = 4.2$; RMSEA = .04; CFI = .99. It was the only model which showed a good fit according to Schermelleh-Engel, Moosbrugger and Müller (2003).

2.3.4 Reliability

Cronbach's alpha was calculated with the psych package (Revelle, 2015). For an overview of the reliability of the face cognition tasks see Table 3.

Reliability measures for the face cognition tasks

		ZFCT				
	GFMT short	**Simultaneous matching**	**Sequential matching**	**Face Memory**	**Total**	**CFMT+**
Cronbach's alpha	.65	.64	.71	.74	.83	.89
Trials	40	50	55	39	144	98

Note. GFMT short = Glasgow Face Matching Test short Version; CFMT+ = Cambridge Face Memory Test long version; ZFCT = Zurich Facial Cognition Test. Trials with variance = 0 were deleted.

2.3.5 Group comparisons

For an overview of the performance of the accuracy of the different groups see Figure 5. Bayesian ANCOVAS were conducted with JASP (JASP Team, 2016). Age and gender were controlled by using them as covariates and inclusion as nuisance in the null model. An effect of the group was very strongly supported for the GFMT short (BF_{10} = 44.0) and decisive for the simultaneous matching (BF_{10} = 4.713e +6) and the sequential lineups (BF_{10} = 118.3). The results for the face memory part (BF_{10} = 1) and the CFMT+ (BF_{01} = 1.8) were ambiguous and neither the null hypothesis nor the group hypothesis was supported clearly.

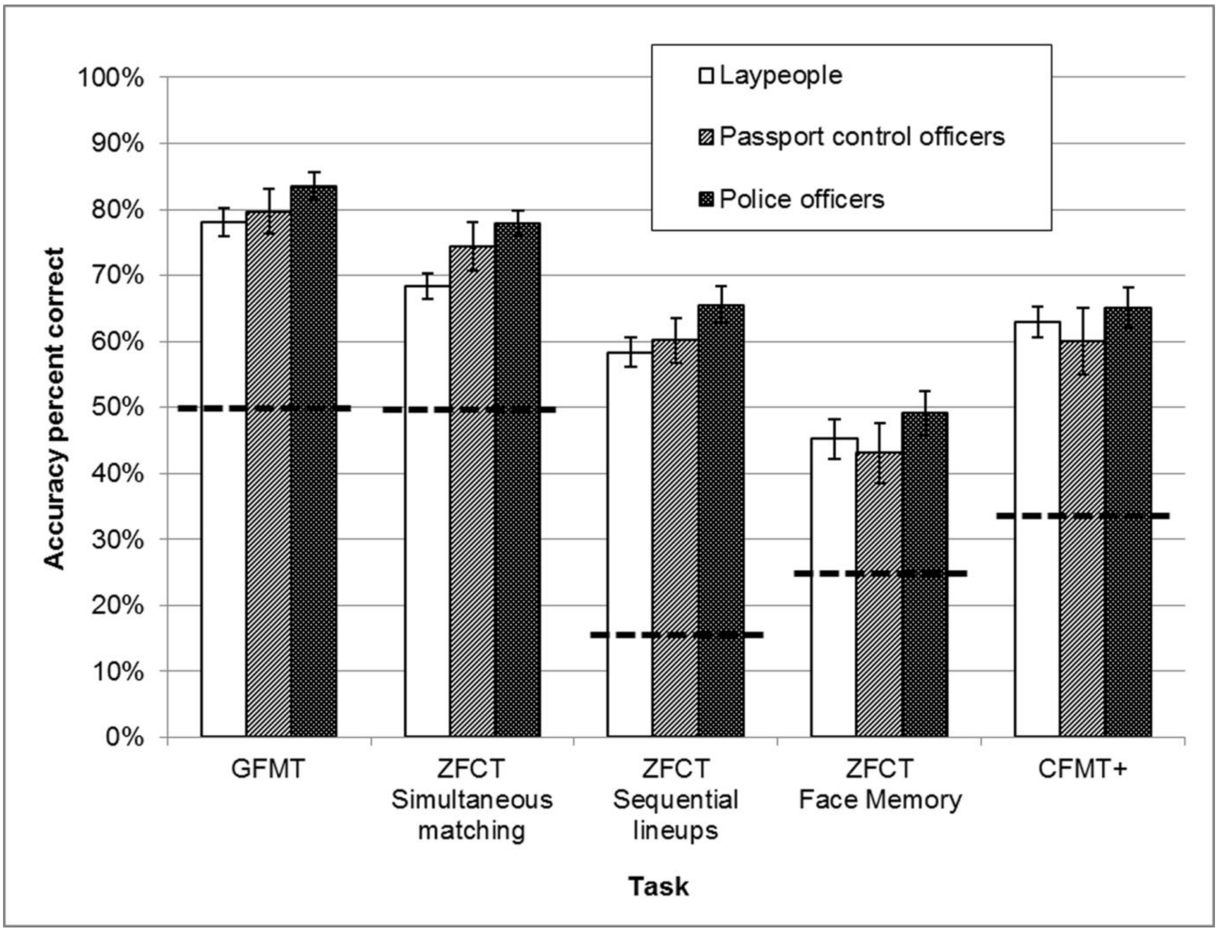

Figure 5. GFMT = Glasgow Face Matching Test short version; ZFCT = Zurich Facial Cognition Test; CFMT+ = Cambridge Face Memory Test long form.
Accuracy of the different groups. Error bars represent 95% confidence intervals. The dashed lines depict the guessing probabilities of the tasks.

Tukey post hoc tests showed that police officers performed better than laypeople in the GFMT short [laypeople 78.1%, police officers 83.6%; $t(167) = -3.6$, $p < .01$; $BF_{10} = 41.2$], the simultaneous matching [laypeople 68.4%, police officers 77.8%; $t(167) = -6.7$, $p < .001$; $BF_{10} = 2.910e+7$] and the sequential lineups [laypeople 58.4%, police officers 65.6%; $t(167) = -4.1$, $p < .001$; $BF_{10} = 179.6$]. There were no significant differences between police officers and laypeople in the face memory part and the CFMT+. The evidence in favor of the null hypothesis was substantial for the CFMT+ and anecdotal for the face memory part. Passport control officers and laypeople performed equally in all face cognition tasks except the simultaneous matching [laypeople 68.4%, passport control officers 74.5%; $t(128) = -3.1$, $p < .01$; $BF_{10} = 9.3$]. There were no significant differences between passport control officers and police officers in any of the face cognition tasks and the BF's were all ambiguous. For an overview of all Bayes factors from the t-tests see table 4.

Table 4

Bayes t-tests for independent samples

		GFMT short	ZFCT Simultaneous matching	ZFCT Sequential lineups	ZFCT Face Memory	CFMT+
PO vs. LP	BF_{10}	41.2	2.910e +7	179.6	---	---
	BF_{01}	---	---	---	1.5	3.3
PO vs. PPO	BF_{10}	1.2	---	1.7	1.4	---
	BF_{01}	---	1.1	---	---	1.0
LP vs. PPO	BF_{10}	---	9.3	---	---	---
	BF_{01}	3.5	---	3.5	3.6	2.5

Note. GFMT short = Glasgow Face Matching Test short Version; CFMT+ = Cambridge Face Memory Test long version; ZFCT = Zurich Facial Cognition Test.
PO = Police officers; LP = Laypeople; PPO = Passport control officers
BF_{10} = Bayes factor in favor of the H_1; BF_{01} = Bayes factor in favor of the H_0

2.4 Discussion

2.4.1 Psychometric quality of the Zurich Facial cognition test

Validity. Performance in the Zurich Facial Cognition Test shows a high relationship to known tests in face cognition. Compared to the CFMT+ and the GFMT short, which measure only one aspect of face cognition, either face perception or face memory, the ZFCT covers both factors. As shown by a confirmatory factor analysis, the sequential lineup part of the ZFCT measures both face perception and face memory. The results could also replicate the finding of Herzmann et al. (2008) that face perception and face memory are two different factors and should be considered separately. In contrast to the test battery of Herzmann et al. (2008), real faces are used, which is closer to the demands of face cognition in applied settings. The ZFCT is also more realistic than the CFMT+ and therefore has a higher face validity, which is important for the acceptance of participants when a test is deployed for personnel selection.

Reliability of the scales. The whole test demonstrated a high reliability. The single test parts have only an acceptable alpha. As Cronbach's alpha is a measure for the internal consistency of a test, the low alpha of the test parts alone could be due to the fact that the test pictures are highly individualized and lack standardization. Nevertheless it has to be suggested to use the ZFCT in its whole form to evaluate individual performance, although for example for the selection of passport officers, face memory is less important than face perception.

Gender and age. Gender had no effect on performance in the ZFCT. Women and men, therefore, have equal chances to achieve the required performance. There was a negative relation between age and performance in the memory part of the ZFCT and the CFMT. Although in comparison to other cognitive abilities face memory seems to peak late (Germine, Duchaine, & Nakayama, 2011), a decline with age has been found for memory-related face cognition tasks (Boutet & Faubert, 2006; Searcy, Bartlett, & Memon, 1999). Meta-analytical results also showed an own-age bias with a superior performance in recognition for faces of the participant's own age group compared to other age groups (Rhodes & Anastasi, 2012). As more faces in the ZFCT and all faces in the CFMT are younger adults, both effects can have had an influence on performance in the face memory part of the test.

2.4.2 Group differences

Police officers were better not only in the simultaneous matching and the sequential lineup part of the Zurich Facial Cognition Test but also in the Glasgow face matching test. This finding contradicts existing literature comparing laypeople with security personnel (Burton et al., 1999; White, Kemp, Jenkins, Matheson, et al., 2014; Frey, C. I., 2008). Performance of the passport officers was in between lay people and police officers. Passport control officers demonstrated a significantly better performance compared to lay people, but only in the simultaneous matching part, which corresponds closest to their daily work. There were no differences between passport officers and laypeople in other tasks and no significant differences between passport officers and police officers in any of the tasks. These findings of group differences were unexpected and several possible reasons could account for these results. One possible explanation is a motivation effect. Police officers could have been aware of the importance of face cognition in their job and therefore could have taken the task more seriously than laypeople. Maybe they performed the tests more conscientiously because they are aware of the consequences of a false decision. Furthermore, police officers are exposed to a competitive corporate culture, which could have enhanced their motivation additionally. Contradicting to this explanation, one would expect an even higher motivation and identification with the task for the passport control officers. However in contrast to the police officers there were no differences between passport control officers and laypeople in the Glasgow Face Matching and the sequential matching part of the ZFCT. Motivation is therefore probably not the only reason for these results. An alternative explanation is that people self-select for jobs in which face cognition plays an important role. It is possible that people who recognize that their performance in face cognition is below average avoid security jobs in which face cognition plays an important role. But this reason should play an even more important role for passport control officers than for police officers. A simple effect of on-the-job practice is also not likely. In this case, passport control officers should show a better performance than police officers. One of the differences between the study of White, Kemp, Jenkins, Matheson, et al. (2014) and the Zurich Facial Cognition Test is, that White and colleagues only used Caucasian faces whereas the ZFCT also consists of other race faces. In addition, the ZFCT is more realistic and less standardized than tasks in other studies regarding color, disguise or changes of the viewpoint. However, these differences cannot explain the better performance of the police officers compared to laypeople in the GFMT short.

Most likely police officers and passport control officers in this study used strategies laypeople don't know. These strategies seem to be effective for tasks closer related to face perception but not for tasks related more strongly to face memory. This is in accordance with the fact that face perception seems trainable to a certain degree (Dowsett & Burton, 2015; White, Kemp, Jenkins, & Burton, 2014) whereas in contrast attempts to train face memory in healthy adults were not successful (Dolzycka et al., 2014). The nature of such possible strategies remains unclear and needs closer examination in future studies. To rule out possible effects of self-selection conclusively, it would be interesting to test police officer or passport control officers novices before the beginning of their education.

2.4.3 Practical implications

Face cognition plays an important role in several tasks accomplished by security personnel. The tasks often take place under conditions which hamper face perception or recognition: People with criminal intent often conceal facial features, photographs of people under current arrest warrants are sometimes of very low quality, and passport pictures can be up to ten years old. These circumstances were taken into account by the test developed and presented in this paper. Training studies have shown that face cognition abilities are hardly trainable (Dolzycka et al., 2014; White, Kemp, Jenkins, & Burton, 2014) and that there are large individual differences in face cognition performance (e.g., Russell et al., 2009). This stresses the importance of a selection tool for occupations in the security domain which measures corresponding face cognition ability demands.

The test has a rather high difficulty level and might therefore also be used to assess super-recognizers for positions with a strong emphasis on face cognition, such as police officers employed in undercover investigations, observations, or the image search domain. According to Russell et al. (2009) the cut-off value for super-recognition can be set at two standard deviations above the mean performance correspondingly to the cut off worth for prosopagnosia on the other end of the scale. Further, the test could be used to test one aspect of the credibility of eye-witness statements in lineup identification processes as Bindemann, Brown, Koyas and Russ (2012) suggest. Witnesses could be tested with face memory tasks to evaluate the likeliness they are able to identify an offender based on his face.

2.4.4 General Conclusion

Overall, the Zurich Facial Cognition Test provides a realistic tool with a high difficulty level to test face cognition requirements for security employees. Possible strategies which improve face perception performance should be examined in further studies.

3 Attention to the ears improves simultaneous face matching accuracy: A simple method to enhance performance in security and forensic settings.

Mirjam Fuhrer[1] and Franziska Hofer[1,2]

[1] University of Zurich, Cognitive Psychology Unit

[2] Kantonspolizei Zürich (Zurich State Police), Research & Development

Submission status

Submitted for publication.

Authors' contributions

MF: Review of the literature, design of the study, development of the test, selection of the stimuli, selection and preparation of the stimuli, data collection, data analysis and interpretation, writing of the manuscript

FH: Supervision and discussion of MF's contributions, revising the manuscript

Acknowledgment

This research was funded by a grant from the Swiss federal office of civil aviation and the Kantonspolizei Zürich (Zurich State Police). Special thanks go to the police officers and passport control officers of the Kantonspolizei Zürich who provided us with insight into their daily work and experience. The authors also thank Amanda Planzer for her assistance with data collection and Klaus Oberauer, Colin Wright and Haley Pruitt for revising the article. Thanks go also to the researchers who provide the images of their databases to the scientific community. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

Abstract

It is well known in forensic anthropology and biometric computer science that the shape of the ears is unique and stable over time and can serve as a reliable visual feature to identify people. Despite this fact, psychological research has paid little attention to the ears so far. This study investigated the ability of human observers to distinguish people based on the shape of their ears as well as the influence of increased attention to the ears on simultaneous face matching performance with unfamiliar faces. Our results show that it is indeed possible to match people if only the ears are available as a source of visual information in frontal portrait photographs. Additionally, accuracy in simultaneous matching of unfamiliar faces was improved in two experiments when attention to the ears was increased. A simple instruction to pay attention to the ears could be helpful for applied settings in which face matching plays an important role, such as passport control along borders.

Keywords:

Simultaneous face matching, external features, ears, biometrics, identity verification

3.1 Introduction

In a simultaneous face matching task, a viewer has to decide whether two simultaneously depicted pictures of faces show the same person or two different people. The task is of great importance for identity verification by photo ID's in security and forensic applications. One of the most typical examples for this task is passport control along borders.

Previous research has consistently shown that human face matching performance is far from perfect for unfamiliar faces with pictures (e.g., Henderson, Bruce, & Burton, 2001) as well as for matching of a live person with photographs (Davis & Valentine, 2009; Kemp et al., 1997; Megreya & Burton, 2008). Nevertheless, the face remains the most common feature to identify a person in an efficient way and without specialized field specific technological equipment.

While unfamiliar face matching is already difficult when conditions are standardized, applied settings make the task even more challenging. Passport pictures can be up to ten years old and hair style, hair color or facial hair are often subject to changes over time. In addition to these challenges, there are large individual differences in face matching ability, ranging from perfect to guessing probability level (Henderson et al., 1999; Lee, Vast, & Butavicius, 2006). Mere practice does not seem to improve performance. Examining passport control officers, White, Kemp, Jenkins, Matheson, et al. (2014) found no correlation between employment duration and face matching accuracy and no difference in performance between passport control officers and laypeople. In contrast to face memory accuracy which seems hardly trainable (e.g., Dolzycka et al., 2014), research has demonstrated that face matching performance could be improved through direct trial-by-trial feedback (Alenezi & Bindemann, 2013; White, Kemp, Jenkins, & Burton, 2014). Furthermore, other research findings revealed that after participants worked together in pairs, individual performance was improved after the collaboration, particularly for participants whose accuracy was initially lower (Dowsett & Burton, 2015). However, the improvements are small compared to the initial individual differences. In a study with police officers, passport control officers and laypeople, we found differences in performance between police officers and laypeople in the short version of the Glasgow Face Matching Test (Burton et al., 2010), as well as in the simultaneous matching and the sequential lineup tasks of the Zurich Facial Cognition Test (Fuhrer & Hofer, 2016). Additionally, passport control officers demonstrated better performance than laypeople in the simultaneous matching subtest of the Zurich Facial Cognition Test. Aside from possible motivational differences between security personnel and laypeople, these results might be attributable to strategies that are useful for face matching tasks. Asking the police officers for strategies they used in the task, some reported to pay attention to the shape of the ears if they are uncertain about the decision.

In addition to this anecdotal evidence, the visual appearance of the ears has been used for a long time for person identification in forensic anthropology (Imhofer, 1906). Some studies examined the uniqueness of ears and provided empirical evidence that ears are individual and therefore represent a reliable way to differentiate between people (Purkait, 2016; Purkait & Singh, 2008). The most prominent examination is by Iannarelli (1989) who found no indistinguishable ears in a sample of over 10’000 ears. He reports that the shape of the ears is stable over the life span as the growth of the ear is highly linear. Further, Yan and Bowyer (2005) found in a sample of 404 people that about 90% of people have symmetric ears.

Apart from forensic anthropology, the ears have also been proposed as biometrics by computer scientists. Several algorithms have been developed using the 2D or 3D shape of the ears to identify people with biometric technology (e.g., Burge & Burger, 1996; Hurley, Nixon, & Carter, 2005; Kumar & Wu, 2012). Recently, a smartphone application was developed to establish patient's identities using a smartphone camera image of their ear. The application is intended to manage medical records in less developed countries in which conventional forms of patient identification are lacking or poorly maintained (Bargal et al., 2015).

Studies about human ability to distinguish ears on a perceptual level are rare. Hoogstrate, Van Den Heuvel and Huyben (2001) tested whether 6 experts and 6 laymen were able to identify people on the basis of the shape of their ears from surveillance camera images. They came to the conclusion that this was indeed possible. Experts made no false identifications when they indicated to have enough information to make a safe decision. Therefore, it seems possible that humans are able to differentiate between people based on the visual appearance of their ears. The above mentioned paper does not describe the exact procedure and it remains unclear how long or often the videos could be viewed. Further, there is no information on the viewpoint of the ears and it remains unclear whether the task could be similarly performed if only frontal views of the ears were available.

Despite the well-known use of the ears for person identification in forensic science or for biometric technology, psychological research has paid no attention to the potential of ears for personal identification purposes. We assume, that ears are processed like other visual objects and that they are not automatically used in the processing of faces because we have little experience in processing ears. Faces attract attention automatically as they are very important sources of information and convey a complex variety of important messages used by humans in social contexts, such as in relation to feelings, age, sex or the other person's focus of attention (Bruce & Young, 2012). Already newborn infants show a preference for face-like images over scrambled versions of pictorial patterns (Goren, Sarty, & Wu, 1975). McKone and Robbins (2011) review the discussion in research if faces are special compared to other visual objects and come to the conclusion that there is now good evidence, that face processing is different from the processing of other visual stimuli. There are unique visual mechanisms for the processing of information in faces compared to other objects. Holistic or configural processing is limited to faces and there are dissociable cortical regions dedicated to processing faces. In line with this, twin studies demonstrated heritability of face recognition ability independent of object recognition and general cognitive abilities (Wilmer et al., 2010).

People naturally seem to pay little attention to the ears. According to Bruce and Young (2012), ears tend to be an unremarkable aspect of human facial appearance, partly because our interactions are often face to face. From the frontal perspective of a face, ear shape is not fully visible and ears are often concealed by hair.

The ears have never been considered particularly in psychological research, but they appear sometimes subsumed under the category of *external features*. The features of faces can be separated between internal features which refer to the central region of a face (e.g., eyes, nose and mouth) and external features (hair, face outline, chin, facial hair). Some studies categorize the ears with external features (e.g., Fletcher, Butavicius, & Lee, 2008), others don't mention the ears in their specification of external features (e.g., Bonner, Burton, & Bruce, 2003).

The distinction between internal and external features has been used to explain differences in processing of familiar and unfamiliar faces. In contrast to unfamiliar faces, memory and matching of familiar faces is very accurate even under difficult conditions such as poor quality video material (Burton et al., 1999). This was attributed in some studies to the observation, that matching and recognition of familiar faces is mainly based on internal features (Ellis, Shepherd, & Davies, 1979; Henderson et al., 1999), whereas representation of unfamiliar faces are found to be dominated by external features, especially hairstyle (Ellis et al., 1979; Frowd, Bruce, McIntyre, & Hancock, 2007). Likewise, in the process of familiarization with new faces, internal face features gain increasing importance (Bonner et al., 2003) and the internal-feature advantage has been proposed as an index of familiarity (Clutterbuck & Johnston, 2002, 2005; Osborne & Stevenage, 2008).

External features are in these studies often generally considered as unreliable information, because changes of most external features are likely as time goes by (Fletcher et al., 2008). Face outline and chin are influenced by increases or decreases in weight and hair style and facial hair is likely to be changed as well. Although in contrast to the external features, the shape of the ears is stable over the life span (Iannarelli, 1989), in these reflections, the ears have not been considered specifically. This disregards the potential valuable contribution ears could have in person identification contexts.

Under certain circumstances, the ears could even be a more reliable source of information than the face itself, in particular in conditions in which face matching is difficult. As previously mentioned, this is the case when factors occur like differences in emotional expression (Bruce, 1982; Henderson et al., 1999), viewpoint (Liu & Chaudhuri, 2002; Meinhardt-Injac et al., 2009), lighting (Hill & Bruce, 1996) or age (Megreya et al., 2013). Performance in face matching is also worse with other ethnicities (Megreya, White, & Burton, 2011). Apart from stability over time, the ears have an additional advantage compared to the face. In contrast to the appearance of the internal features, which are highly influenced by expressions, the ears are independent from the movements of facial expressions. Given the error-prone accuracy in face matching tasks and their implication for applied settings like passport control, the instruction to use the ears as an additional source of information could therefore be a useful method and a very easy tool to improve performance, especially in cases when some of the before mentioned factors occur and practitioners are uncertain with a decision based exclusively on the face.

The present study is comprised of two experiments. The objective of the first experiment was to investigate performance in a simultaneous matching task depicting only the ears of two people presented in frontal view. In addition, matching performance with stimuli of pixelated faces and normal ears (*only ears*), faces without ears (*no ears)* and intact faces (*intact*) containing both features were compared to investigate the impact of the ears and the face itself (without ears) as different sources of information. In the second experiment, the influence of a mere instruction to pay attention to the ears on face matching performance was examined with intact faces. Eye-tracking was applied in experiment 2 to monitor the initial attention to the ears prior to instruction as well as the later compliance with the instruction.

3.2 Experiment 1

3.2.1 Method

Subjects. Forty participants recruited from university (9 male) with a mean age of 33.1 years (*SD* = 14.7) volunteered to participate. They received either course credits or a small fee and provided informed consent.

Stimuli. Face photographs were retrieved from different databases: The Glasgow Unfamiliar Face Database (Burton et al., 2010), the FEI Database (De Oliveira Junior & Thomaz, 2006), the FERET database (Phillips et al., 2000, 1998), the Minear & Park Face Database (Minear & Park, 2004), the PICS (University of Stirling, n.d.), the FG-NET ageing database, which is available upon request (for a review of research using the database see Panis & Lanitis, 2014), and some faces were retrieved from the Interpol database of wanted people (Interpol, n.d.). Additionally, two unpublished face databases were used: The Taiwanese National Chung Cheng University Face Database from Jisien Jang, and a Database from Corinne Koller and Sarah Chiller-Glaus, University of Zurich. Altogether, 336 Photographs of 252 different people were chosen. All faces were depicted in frontal view with ears clearly visible and not obscured by hair. Each photograph was used only in one trial. Backgrounds and visible clothing were removed and the faces were cropped neatly around the external features of the face using graphical software (see Figure 1 intact trial type). A total of 168 simultaneous matching trials were constructed, in which two faces were presented side by side at the same time. The faces were resized to a height of 650 pixels and positioned on a 1680 x 1050 pixels white background. They were aligned in such a way that the pupils were at a height of 475 pixels and the horizontal distance between the bridge of the nose of the two images was 840 pixels. Half of trials were matches, containing two images of the same person. In order to increase the difficulty level, 54 of the matching trials were constructed of two photographs which were not taken on the same day, and 41 of these trials involved documented time differences of more than a year between the two pictures. The other 30 match trials, which were constructed of photographs from the same day, differed in expression or were taken with different cameras. The other half of trials were mismatches, depicting two different people. For mismatches, the first author chose pictures from subjectively similar looking people to increase difficulty level. Fifty-six of the match trials were constructed of Caucasian faces, 28 of Asian faces, and the same proportion was used for construction of mismatch trials. The majority of the trials comprised of male faces, except 12 match trials (three Caucasian, nine Asian) and seven mismatch trials (two Caucasian, five Asian) with female faces.

In a second step, two additional types of the 168 trials using the same faces were created to obtain three different conditions (see Figure 1). One condition consisted of unchanged faces (*intact*), one of faces without ears (*no ears*) and one of ears without face information (*only ears*). For the *no ears*-condition, ears were painted over with white color neatly along the face line using Photoshop CS4. For the *only ears*-condition, the faces were pixelated so that only the ears were depicted in full sharpness. In order to avoid the possibility that any diagnostic information could be gained from the appearance of the pixelated faces in the *only ears*-condition, only one of the faces of each trial was pixelated and then placed over the other face (see figure 1 for the three different trial types).

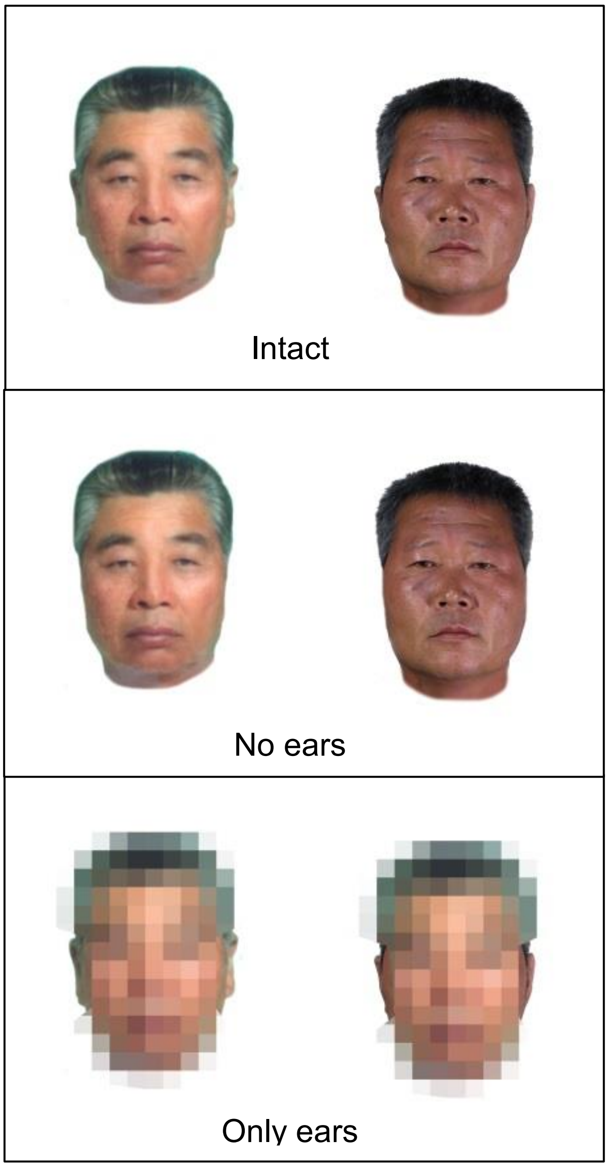

Figure 1. Example of the different trial types used in experiment 1, depicting an identity mismatch

To prevent effects based on the difficulty of the trials or differences in similarity of the ears, the trials were divided into four sets of 42 trials, each containing 14 Caucasian matches, seven Asian matches and the same proportion of mismatches. These four sets of face pairs were rotated over the four conditions so that each set appeared equally often in each condition, leading to four different versions of the experiment (see Table 1). The four sets were equated as much as possible with regard to gender and database.

Table 1
Face set per trial type in experiment versions

Experiment version	Block 1 Pre instruction	Block 2 Post instruction		
	Intact	No ears	Only ears	Intact
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

Note. A, B, C and D were four different face sets used in different trial types (*intact*, *no ears* and *only ears*) in experiment versions 1,2,3 and 4. Each face set consisted of 1/3 Asian and 2/3 Caucasian faces and half of the trials were mismatches. Order of trial presentation was randomized within each block.

Procedure. Participants were randomly assigned to one of the four experiment versions. In the beginning, they were presented with a written explanation of the task on the screen and instructed to press button A on the computer keyboard, if they believed that the two photographs depict the same person and button L if they thought that the photographs come from different people. To highlight the buttons, A was associated with a green sticker and L with a red sticker. Trials were presented for up to twenty seconds, or until participants gave an answer. After the 20 seconds elapsed, the faces disappeared from the screen. Participants had to press A or L to start the next trial. In the first test block, participants were presented with one set of 42 *intact*-trials to obtain a baseline of their face matching performance. After this block participants were presented with a detailed written instruction, which informed them that ears are as unique as fingerprints and therefore could be valuable to decide whether two pictures show the same person or two different people. The instruction included several examples with pairs of faces with visible ears. After this instruction, participants solved 126 trials of three sets in randomized order, one set of *intact*-trials, one set with *no ears*-

trials and one set with *only ears*-trials. For each participant, the same faces were never shown in different conditions and the presentation of all faces was counterbalanced across participants with the different experiment versions.

3.2.2 Results

Data analyses. As the traditional use of p-values in science has been criticized and there are several shortcomings associated with classical null-hypothesis testing (e.g., Wagenmakers, 2007; Wasserstein & Lazar, 2016), Bayes factors are calculated and reported in this paper in addition to the p-values (for a short introduction into Bayes statistics see e.g. Jarosz & Wiley, 2014). A Bayes factor is a ratio that compares the likelihood of the data under two different hypotheses, typically the null and the alternative hypothesis. The Bayes Factor scale is continuous, the value reflects how many times more likely the data are under a hypothesis in the numerator relative to the hypothesis in the denominator. A BF_{10} larger than 1 reflects evidence favoring the alternative hypothesis. A BF_{10} below 1 favors the null hypothesis, in this case the BF_{01} is reported and expresses how many times more likely the null hypothesis is compared to the alternative hypothesis. For example, a BF_{10} of 20 means that the data are twenty times more likely under the alternative hypothesis than under the null hypothesis. To interpret the size of the Bayes Factors, a verbal description based on Jeffreys (2000) is additionally provided. Bayesian analyses were calculated using JASP (JASP Team, 2016) and the default priors were used. Classical analyses were calculated with R (R Core Team, 2014).

General performance. Matching performance was calculated using signal detection theory (Green & Swets, 1966). To calculate the accuracy measure d', the z-transformed false alarm rate is subtracted from the z-transformed hit rate. A hit was scored, when a mismatch trial was correctly classified as a mismatch. A false alarm was scored if a match trial was falsely classified as mismatch. The hit rate was the proportion of hits over all mismatch trials and the false alarm rate was the proportion of false alarms over all match trials. For an overview of the performance in the four conditions see Figure 2.

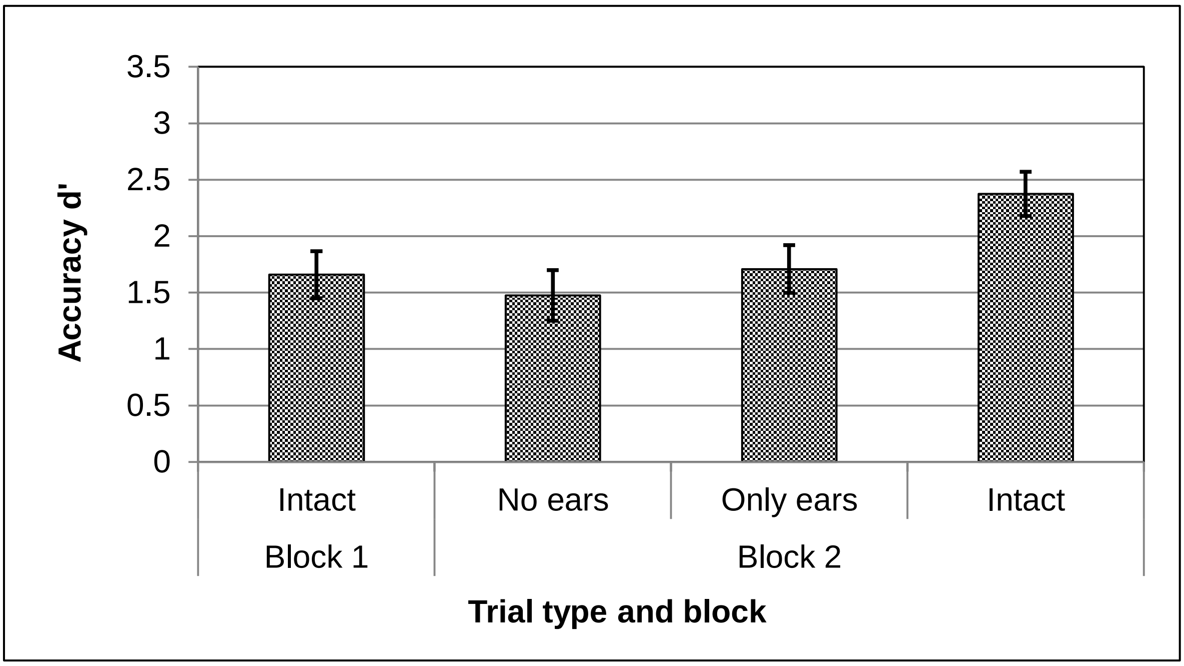

Figure 2. Mean accuracy in experiment 1 with trial types before (block 1) and after (block 2) instruction. Error bars represent 95% confidence intervals.

A repeated measures ANOVA yielded a significant effect of the repeated factor with a decisive Bayes Factor supporting the alternative hypothesis, $F\ (3, 117) = 19.89$, $p < .001$; BF_{10} = 9.091e+7. Performance in the *only ears*-trials was tested against chance level (d' = 0) to evaluate, if participants were able to match the face pairs based only on ears as visual information. Performance was clearly above chance level $t(39) = 15.86$, $p < .001$; BF_{10} = 3.751e+15. To assess if there was a general improvement in performance after the instruction which cannot be attributed to attention to the ears, performance in *intact* trials before the instruction was compared to performance in *no-ears* trials after the instruction. There was no significant difference. The Bayes Factor reflected an ambiguous result, supporting neither the null nor the alternative hypothesis clearly $t(39) = 1.7$, $p = .09$; $BF_{10} = 0.67$, $BF_{01} = 1.49$. Accuracy in *no-ears* trials was then compared to *intact* trials after instruction to test the impact of the presence or absence of the ears in two conditions after the instruction. Accuracy in *no-ears* trials was significantly lower than in the *intact*-trials after instruction $t(39) = -5.44$, $p < .001$; $BF_{10} = 5917.57$. Performance in *intact* trials was compared before and after the instruction to evaluate differences based on the phase of the presentation. Performance with intact face pairs was significantly lower before than after the instruction $t(39) = -6.99$, $p < .001$; $BF_{10} = 617418.57$

3.2.3 Discussion

The main finding of this experiment reveal that the matching of people based on the shape of their ears is possible.

Performance in the *only ears*-trials was clearly above chance level. This shows that human observers are able to perform simultaneous matching of people based on the shape of their ears. The perspective of frontal face portraits in which ears are only visible at the side of the face is sufficient for this purpose. This is important to note as pictures in identity documents are frontal view photographs. Performance with *only ears*-trials was on a comparable level to performance for *no ears*-trials or for *intact*-trials in block 1 before the instruction. It is possible that participants did not pay attention to the ears and therefore processed *intact*-trials prior to the instruction like *no-ears* trials. This would be in line with the finding, that external features in general don't receive much attention, whereas the internal features are fixated a lot (Fletcher et al., 2008; Van Belle, Ramon, Lefèvre, Rossion, & others, 2010). More surprising, however, is that accuracy with *only-ears* trials was similar to the other two conditions in which only faces were available respective used. This result show, that if face matching processes are made difficult by factors which are quite common in applied settings, such as age differences or other ethnicities, the ears might be a source of information which is at least as reliable as the faces themselves.

When participants were instructed to pay attention to the ears, they achieved a better performance in *intact*-trials after than before this instruction. A practice effect with the faces alone seems rather unlikely to be the only reason for this effect. In this case, one would expect no difference in performance between *no-ears* and *intact* trials, as both were presented in random order after instruction. However, a practice effect cannot be clearly ruled out based on experiment 1. As stimuli of different trial types in experiment 1 were presented in random order within block 2, it is also uncertain whether mechanisms of implicit learning caused by the *only ears*-trials improved performance of *intact*-trials in addition to the instruction to pay attention to the ears. Thus, a general learning effect cannot be excluded in experiment 1. The experience participants gained from trials with *only ears*-trials might have enhanced their trust in the strategy more than a simple verbal instruction would do. Only *intact* trials were used in experiment 2 to rule out the possibility of implicit learning through the trials with only ears and in order to be able to examine the sole influence of an instruction. As mentioned, performance was not different between *intact* trials before instruction and *no-ears* trials. This supports the assumption that participants did not use the shape of the ears as a source of information to perform face matching in block 1. However, as the amount of improvement in *intact* trials varied between participants, it is still possible that some used the strategy already in block 1. Improvements in performance due to the strategy would then be underestimated. To monitor possible influences of attention to the ears before a corresponding instruction, eye-tracking was applied in experiment 2.

3.3 Experiment 2

In this experiment we examined the influence of a simple instruction to pay attention to the ears. Additionally, initial attention to the ears and later compliance with the instruction was assessed using eye tracking.

3.3.1 Method

Subjects. Sixty Participants (9 male) recruited from university with a mean age of 29.0 years (*SD* = 8.6) volunteered to participate in this experiment either for a small fee or course credit. All participants reported normal or with the aid of contact lenses corrected-to-normal vision and had not participated in experiment one.

Stimuli. All 168 trials from experiment 1 were used in the *intact*-trial version.

Apparatus. Eye movements were registered with SMI (Senso Motoric Instruments) iView X RED500 tracking system. Data of the left eye were recorded with a sampling rate of 500 Hz, a spatial resolution of 0.03 deg, and a gaze position accuracy of < 0.4 deg. Stimuli were presented on a computer screen with a resolution of 1680 x 1050 pixels. Head movement was restricted by a chin rest. In the beginning of the experiment, a 5-point calibration procedure was performed using iView X (Version 2.8; SensoMotoric Instruments, 2012) software to establish the relationship between the position of the eye in the camera view and the gaze point of the participant.

Procedure. Participants were randomly assigned either to an experimental or a control group. After they took a place in front of the screen and calibration was successfully finished, the same simultaneous matching task as in experiment 1 was conducted (168 trials). In contrast to experiment 1, only *intact*-trials were used. Additionally, before every trial a fixation cross was presented for 500 ms in the center of the screen. The first half of the trials (84) were presented after a written instruction of the simultaneous matching task in both groups. After the first block, participants of the experimental group received the same instruction as in experiment 1, recommending to use the ears as source of information additional to the faces. In contrast, participants of the control group received only the information that they had solved half of the trials. Then, the second half of the trials was presented (block 2). Facial stimuli were counterbalanced across participants within experimental groups; half of participants in every group were presented with face sets A and B in block 1 and face sets C and D in block 2, and the other half received the reverse assignment of face sets to blocks.

3.3.2 Results

Eye fixations. Faces and ears were defined as different areas of interest (AOI) using a freehand tool from BeGaze (Version 3.3; SensoMotoric Instruments, 2013). See Figure 3 for an example trial with AOIs. The AOI borders only served the purpose of data analysis and were not visible during the experiment.

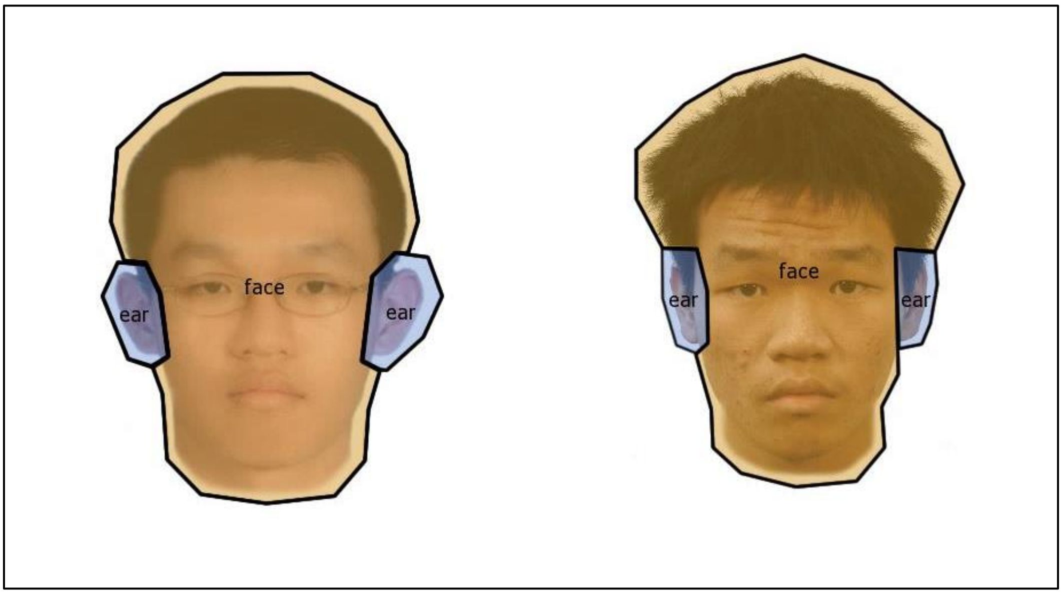

Figure 3. Example for locations of areas of interest (AOI) on a mismatch trial. This AOI borders were invisible for participants during the experiment. The two faces and four ears were grouped together in analysis. The surrounding area of the screen was defined as white space.

The two faces and four ears of each face pair were combined to two AOI groups (i.e., a face and an ear group) and the surrounding area of the screen was defined as white space. For each person, the sum of the fixation duration on each AOI group (face, ear) was calculated over all trials of one block and then divided by the total fixation duration of the person per block. The relative fixation duration on AOI groups per block are depicted in Figure 4.

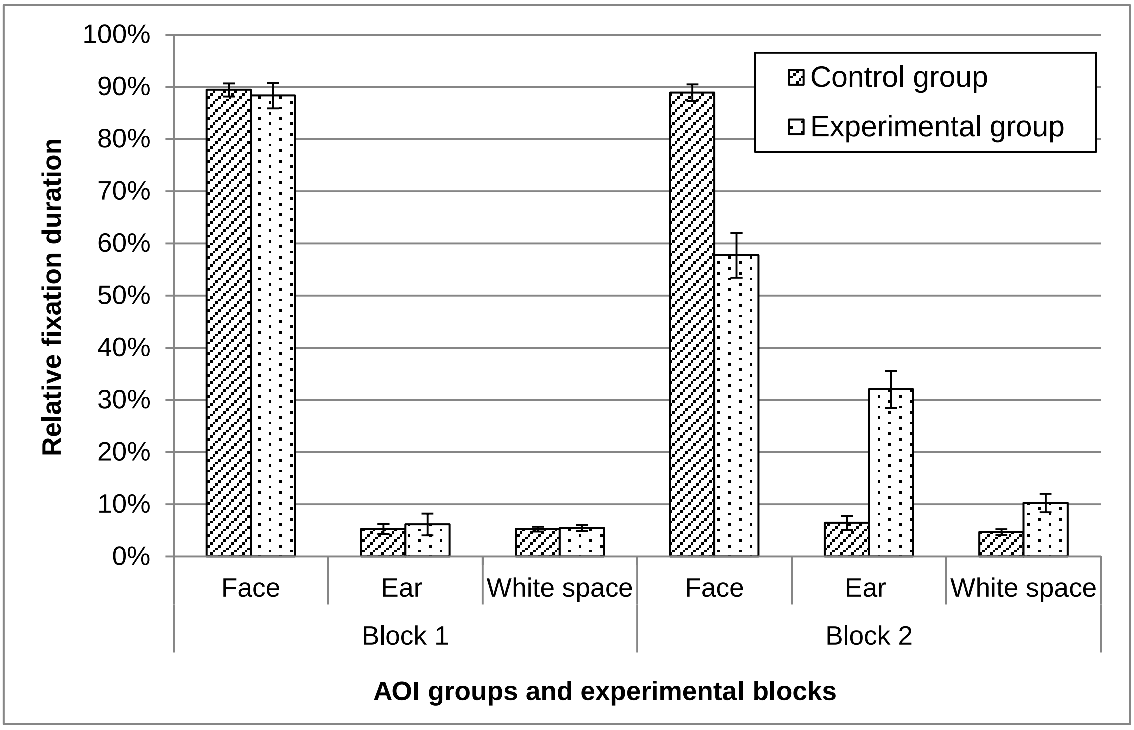

Figure 4. Proportion of fixation time per AOI group and experimental blocks before and after instruction for experimental and control group. Error bars represent 95% confidence intervals.

Two 2 x 2 mixed ANOVAs with relative fixation time on each AOI group as the dependent variable, group as between subject factor and block as within subject factor were conducted. There were significant interactions between group and block for the AOI group ear $F(1, 58) = 68.2$, $p < .001$ and the AOI group face $F(1, 58) = 67.5$, $p < .001$. Comparing the hypothesis that there is an interaction with a null model including only the main effects of group, and block and the subject random effect, decisive evidence for the interaction for ear (BF_{10} = 3.981e+10) as well as the face (BF_{10} = 4.457e+10) AOI group was obtained from data. The relative fixation durations on ears and faces were equal in the control and the experimental group in block 1; there were no significant differences observed for the AOI group ear $t(58) = -0.5$, $p = .60$, $BF_{01} = 3.4$ and for the AOI group face $t(58) = 0.5$, $p = .59$, $BF_{01} = 3.4$. A comparison between blocks for the experimental group showed, that after the instruction (block 2), relative fixation duration shifted to a significantly higher amount on the AOI group ears $t(29) = -8.8$, $p < .001$, $BF_{10} = 1.080e+7$ and a lower amount of fixation time spent on AOI group face $t(29) = 8.6$, $p < .001$, $BF_{10} = 5.870e+6$. In contrast, there were no significant differences between blocks for control group participants for the ear AOI group $t(29) = -1.8$, $p = .07$ and the face AOI group $t(29) = 0.7$, $p = .48$. The corresponding Bayes factors reflect an ambiguous result neither supporting the null nor the alternative hypothesis clearly for the ears ($BF_{01} = 1.1$) and substantial evidence for the null hypothesis for the face AOI group ($BF_{01} = 4.1$).

General performance. Performance was measured by d' the same way as in experiment 1. A 2 x 2 mixed design ANOVA was conducted with block as the within subject factor and group as the between subject factor. Most importantly there was a significant interaction between group and block $F(1, 58) = 7.3$, $p < .01$, indicating that the increase in performance from block 1 to block 2 was larger for the experimental than for the control group. Comparing the likelihood of the data under the assumption that there is an interaction with a null model including only the main effects of group, block and the subject random effect, the alternative hypothesis was decisively supported ($BF_{10} = 4.088e+10$). For a graphical illustration of the interaction see figure 5. The initial performance in block 1 was not significantly different between experimental and control group $t(58) = -0.9$, $p = .38$, with a Bayes Factor weakly supporting the null hypothesis of no differences between groups ($BF_{01} = 2.7$). After the instruction in block 2, there was a significant difference in performance between the two groups $t(58) = -3.2$, $p < .01$, BF_{10} = 14.3. Within the control group, a paired samples t-test revealed a significant difference between block 1 and block 2 $t(29) = -2.3$, $p <.05$. However, the Bayes Factor showed only anecdotal evidence for the alternative hypothesis compared to the null hypothesis (BF_{10} = 1.8). Within the experimental group, the difference between block 1 and block 2 was highly significant, and supported by a decisive Bayes Factor in favor of the alternative hypothesis $t(29) = -5.3$, $p < .001$, $BF_{10} = 1848$.

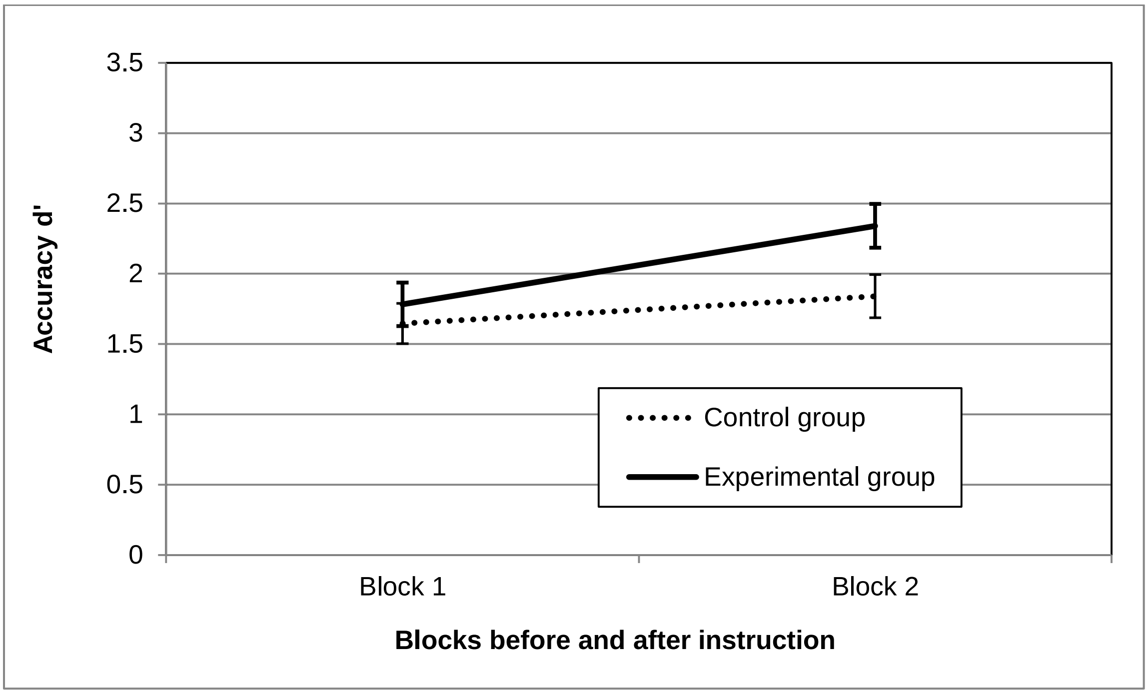

Figure 5. Mean accuracy for experimental and control group. Error bars represent 95% confidence intervals.

3.3.3 Discussion

An instruction to pay attention to the ears was effective to increase performance in simultaneous face matching. As eye tracking revealed, after the instruction participants were able to pay attention to the ears without neglecting the face.

In block 1 prior to the instruction, the relative fixation duration on ears and faces was equal in experimental and control group. Participants spent most of the time fixating on the face and scarcely any time looking at the ears. This is in line with the observation that people don't pay much attention to the ears (Bruce & Young, 2012) and with our assumption, that ears are generally not used to perform face matching by laypeople. Participants of the control group, who did not receive a special instruction after block 1, maintained this pattern in block 2. In contrast, participants of the experimental group showed an increased amount of relative fixation duration on the ears in block 2, and a decreased amount of fixation to the face. Nevertheless, they still spent more time fixating on the faces than the ears.

The significant interaction between experimental group and block showed that participants in the experimental group improved their performance after the instruction more than participants of the control group. The control group showed a slightly better performance in block 2 than in block 1, but this effect received only weak statistical support. In contrast, the enhancement in performance was plainly evident in the experimental group. This enhancement can clearly be attributed to the instruction to pay attention to the ears, as there were no other differences between experimental and control group.

Additionally, experiment 2 consisted only of intact faces with ears and faces visible in all trials. The improvement in performance was generally weaker than in experiment 1. This could be due to two reasons. Firstly, participants in the control group in experiment 2 showed a tendency to demonstrate a practice effect. This practice effect was probably caused by the larger set of trials in experiment 2. Secondly, participants might have paid more attention to the ears in experiment 1 because of the stronger intervention, which consisted in a combination of an instruction as well as of including *only-ears* trials. Having only the ears as available source of information could lead to increased attention to the ears and trust in the strategy in experiment 1 compared to experiment 2. However, a direct comparison between performance in experiment 1 and 2 is not valid as the experimental design differed markedly.

Further studies should compare the two methods directly, i.e. whether a simple instruction to pay attention to the ears or a specific training with *only ear*s-trials has a stronger impact on face matching performance and whether such methods or strategies have long-lasting effects on performance. Furthermore, improvements in performance could be especially high for participants with lower face matching ability. Because their performance is initially low, there is more room for improvement. Likewise, it is possible that participants who find it difficult to match faces are more motivated to use a strategy in which face processing is not necessary. Participants with prosopagnosia would be an interesting subsample for further studies addressing these questions.

3.4 General discussion

In this study we report two experiments investigating the influence of increased attention to the ears in simultaneous matching tasks with unfamiliar faces. We found an increased performance in conditions in which faces and ears are both used as sources of information compared to conditions in which only the face (or only the ears) are used. Improvements in performance with intact faces were found after an instruction and trials with stimuli containing only ears in experiment 1. An instruction to pay attention to the ears was sufficient to improve performance in experiment 2, which only involved intact faces. Experience with stimuli containing only ears is therefore not necessary to boost performance. The analysis of eye fixations further revealed that attention to the ears is not common for most participants considered laypeople, but can be increased with a simple instruction.

3.4.1 Implications for further research

There is a tendency in psychological research to classify external features of a face as unreliable because of their instability over time (Fletcher et al., 2008). The result of this study and findings from anthropology show that the ears should not be considered equal to hair, face line and facial hair because the shape of the ears is unique and stable over the life span. In future research focusing on simultaneous face matching, it should therefore be considered that humans are able to match people based on the shape of their ears even when only frontal views of faces are available. Human object recognition and visual processing of ears should be examined more closely in further studies. As we don't pay much attention to the ears in general, it is unlikely that we have a similar memory performance for ears as we have for faces. It is therefore not reasonable to assume that increased attention to the ears would also be helpful in face memory tasks. However, this was not the subject of this study and cannot be assessed based on our findings.

Under the assumption, that faces are processed different than other visual objects, basic research with a clear focus on the understanding of pure face processing mechanisms should consider that visible ears could distort measures, such as face matching ability. As mentioned, face processing is different from the processing of other visual stimuli (McKone & Robbins, 2011). If participants know about the uniqueness of ears and use their appearance in unfamiliar matching tasks, this strategy leads to results which are not clearly attributable to face processing. In these cases, cropping visible ears from face pictures or asking participants in the end of the study if they have paid attention to the ears should be considered. In contrast, research in which results should be generalizable to applied areas, at least some face stimuli with visible ears should be used. This mirrors more closely the information of people's identities we have available in real life. In particular, studies which examine the influence of external and internal features should not disregard the ears. This is especially important, if applied questions such as the influence of head scarfs on face processing are studied (Megreya & Bindemann, 2009; Megreya, Memon, & Havard, 2012; Toseeb, Bryant, & Keeble, 2014; Wang et al., 2015).

Future research could examine more in depth if matching of ears is trainable, and if there are larger effects of expertise for ear than for face matching. Additionally, replications of the reported improvements in face matching accuracy are needed, involving different sets of stimuli. This is notably necessary to evaluate if the use of ears could be particularly helpful to match for instance faces of other ethnicities. Analyses separated by ethnicity in this study were conducted and yielded to mixed results which could not be interpreted meaningfully.

3.4.2 Practical implications

Although it has been known for a long time that face matching of unfamiliar people is error-prone, identity control still relies on documents containing photographs. This may partly be due to the fact that it is an easy and non-invasive way to assess peoples' identities (International Civil Aviation Organization, 2015). As human face matching performance is made difficult by factors commonly occurring in applied settings, strategies to improve and train face matching performance could help practitioners like passport control or police officers and would improve national and international security. Trainings for security personnel could for example work with trial-by trial feedback (Alenezi & Bindemann, 2013; White, Kemp, Jenkins, & Burton, 2014), although the reported improvements are rather small, and it remains to be seen how stable they are over time. Since a simple instruction to pay attention to the ears improved performance of participants in this study, this strategy could help additionally and does not need time-consuming, expensive and complex training procedures.

The most critical problem with the application of the strategy introduced here is that there are no regulations that ears have to be visible in identity document photographs (International Civil Aviation Organization, 2015). Therefore ears are not necessarily visible on passport pictures, as they are often obstructed by hair. It would be simple to ask the live person to show her ears in a passport control process, but this is of no use if they are not depicted in the identity document. Regulations that ears should be visible in IDs would be helpful and easy to introduce, as this would not require changes in normal perspective of the passport photograph equipment or specialized technology.

As a further limitation concerning the generalizability of this study, it has to be mentioned that it remains unclear whether matching of the ears on two frontal portrait photographs is comparable to matching the ear of a passport picture with an ear of an individual physically present. This needs closer examination in future studies conducted in applied settings with live people.

Our experiments provide a starting point, showing that ears should not be disregarded in psychological research. Ears are unique and stable over time and have the potential to improve accuracy in settings in which simultaneous matching of people is required.

4 General discussion

4.1 Summary

In the empirical part of this thesis, two studies were reported which both aim at improving the accuracy in processing the identity of people.

In study 1, we compared laypeople with police and passport control officers and introduced a newly developed test, the Zurich Facial Cognition Test. The results reveal, that security personnel performed better than laypeople in tasks related to face perception, but not in tasks related to face memory. This is in line with research, which aimed at training the face cognition abilities reviewed in the synopsis of this thesis. Attempts to train face memory were unsuccessful (Dolzycka et al., 2014; Malpass et al., 1973; Woodhead et al., 1979), whereas improvements could be achieved in trainings of face perception (Dowsett & Burton, 2015; White, Kemp, Jenkins, & Burton, 2014). This could indicate that security personnel in our study used strategies which are effective to improve face perception but not face memory. As some training procedures conducted by governmental agencies are ineffective (Towler et al., 2014), it is reasonable to assume, that not all security professionals use helpful strategies. This would explain why some studies found no difference between laypeople and security professionals (Burton et al., 1999; White, Kemp, Jenkins, Matheson, et al., 2014).
Considering the characteristics of the test, the ZFCT is suitable to select highly skilled people in face perception and face memory, but could also serve as a pre-occupational tool to prevent the employment of people who are less able to process faces accurately for positions in which face processing is important.

Study 2 was designed to answer two questions. First, we examined, if individuals are able to perform simultaneous matching of people based on the shape of their ears. Second, we investigated if an instruction to pay attention to the ears in addition to the face is effective to improve identity verification in simultaneous matching. Our results show that it is indeed possible to match people if only the ears are available as a source of visual information in frontal portrait photographs. Additionally, accuracy in simultaneous matching of unfamiliar faces was improved in two experiments when attention to the ears was increased. A simple instruction to pay attention to the ears could be helpful for applied settings in which identity verification plays an important role, such as passport control along borders. As the strategy examined in this study is only a simple intervention, which does not require time-consuming and costly training processes, it can easily be implemented for employed security personnel.

4.2 Future directions

The first study is dedicated to personnel selection, and the second study aims at improving the performance in identity verification with a simple strategy, which could easily be implemented for security personnel performing identity verification.

For some tasks, especially those related to face memory, personnel selection seems to be the best method for an optimized security, as attempts to train face memory were not successful (Dolzycka et al., 2014; Malpass et al., 1973; Woodhead et al., 1979). Clearly, much further research is needed to develop successful training techniques and to understand underlying mechanisms. Some promising results were discussed in the synopsis considering face matching or processing of other race faces. The memory for other race faces could be trained as well as perception of faces (e.g., Dowsett & Burton, 2015; Elliott et al., 1973; White, Kemp, Jenkins, & Burton, 2014). Since security personnel at airports especially face a lot of people with other ethnicities, these approaches could be very valuable for aviation and homeland security.

Further, the validity of the test could be compared to measures of job performance related to face processing tasks. However, job performance related to identity verification is difficult to measure. For instance, it is unclear how often people pass borders with documents of another individual successfully. Operational tests with people carrying false documents would be necessary to control for the presence of imposters. Performance in such operational tests, with passport control officers, for example, could be correlated to the results in the ZFCT to assess the validity of the test more closely.

The strategy to pay attention to the ears introduced in study 2 circumvents the large individual differences which were observed in human face processing ability as well as the limited effectiveness in training attempts. We assume that ear processing is different from face processing but this needs to be examined more closely. A substantial body of research, which has compared internal and external features or examined the role of information of single features has not counted the ears as facial features (e.g., Schwaninger et al., 2002). Therefore, more agreement in the exact definition of which parts belong to the face is needed in research. Furthermore, the processing of ears as visual objects is not yet examined in human. The assumption that the introduced intervention is effective for applied settings has to be tested in studies that more closely mirror real life scenarios with actual people, as it is unclear if matching of two pictures of an ear is the same as matching the picture of an ear with a real ear. Our results indicate that laypeople generally don't seem to pay much attention to the ears when they process faces. It is possible, though, that protruding or large ears attract automatic attention, which cannot be observed for average ears. Our study only provides a starting point that attention to the ears might be useful and ears should not be neglected in psychological research concerning face processing.

4.3 Conclusion

The empirical studies presented in this thesis both provide a method which could improve identification in the security domain, taking into account findings of former research about trainability and individual differences in face processing. Future studies have to examine the success of interventions to improve human face processing abilities in field studies. We hope that our research contributes to the attempts to make the world a safer place without increasing errors of incorrect identifications and without increasing restrictions of personal freedom and privacy.

References

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*, 735–753. doi: 10.1002/acp.2968

Bargal, S. A., Wells, A., Chan, C. R., Howes, S., Sclaroff, S., Regan, E., … Gill, C. (2015). Image-based ear biometric smartphone app for patient identification in field settings. In *Proceedings International Conference on Computer Vision Theory and Applications*. Retrieved from http://cs-people.bu.edu/sbargal/visapp_2015.pdf

BeGaze (Version 3.3) [Computer software]. Teltow, Germany: SensoMotoric Instrumens

Benton, A. L., Sivan, A. B., Hamsher, K. deS, Varney, N. R., & Spreen, O. (1994). *Contributions to Neuropsychological Assessment: A Clinical Manual* (2nd edition). New York: Oxford University Press.

Best-Rowden, L., Bisht, S., Klontz, J. C., & Jain, A. K. (2014). Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. In *2014 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1–8). doi: 10.1109/BTAS.2014.6996296

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*, 96–103. doi: 10.1016/j.jarmac.2012.02.001

Bindemann, M., & Burton, A. M. (2009). The role of color in human face detection. *Cognitive Science*, *33*(6), 1144–1156. doi: 10.1111/j.1551-6709.2009.01035.x

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex*, *82*, 48–62. doi: 10.1016/j.cortex.2016.05.003

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLOS ONE*, *11*, e0148148. doi: 10.1371/journal.pone.0148148

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognisers in Action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, *30*, 81–91. doi: 10.1002/acp.3170

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*, 527–536. doi: 10.1080/13506280244000168

Borchard, E. M. (1932). *Convicting the innocent: Sixty-five actual errors of criminal justice.* New York: Yale Univeristy Press.

Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition*, *34*, 854–864. doi: 10.3758/BF03193432

Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: A critical review. *Cognitive Neuropsychology*, *17*, 55–71. doi: 10.1080/026432900380481

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116. doi: 10.1111/j.2044-8295.1982.tb01795.x

Bruce, V. (2011). Applied research in face processing. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook in Face Perception* (pp. 131 – 146). New York: Oxford University Press.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*, 339–360. doi: 10.1037/1076-898X.5.4.339

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Bruce, V., & Young, A. (2012). *Face perception*. New York: Psychology Press.

Burge, M., & Burger, W. (1996). Ear biometrics. In *Biometrics* (pp. 273–285). Springer. Retrieved from http://link.springer.com/chapter/10.1007/0-306-47044-6_13

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*, 1467–1485. doi: 10.1080/17470218.2013.800125

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*, 361–380. doi: 10.1111/j.2044-8295.1990.tb02367.x

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception*. New York: Oxford University Press.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*, 286–291. doi: 10.3758/BRM.42.1.286

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*, 243–248. doi: 10.1111/1467-9280.00144

Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary: Evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia*, *48*, 4057–4092. doi: 10.1016/j.neuropsychologia.2010.09.017

Chang, K., Bowyer, K. W., Sarkar, S., & Victor, B. (2003). Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 1160–1165. doi: 10.1109/TPAMI.2003.1227990

Chiller-Glaus, S. D., Schwaninger, A., & Hofer, F. (2007). Identity verification from photographs in travel documents: The role of display duration and orientation on performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 51,* 1327–1330. doi: 10.1177/154193120705101910

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*, 985–994. doi: 10.1068/p3335

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*, 97–116. doi: 10.1080/09541440340000439

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*, 482–505. doi: 10.1002/acp.1490

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2011). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*, 587–605. doi: 10.3758/s13428-011-0160-2

De Oliveira, L. L. & Thomaz, C. E. (2006). *Captura e Alinhamento de Imagens: Um Banco de Faces Brasileiro*. Undergraduate Technical Report (in portuguese), Department of Electrical Engineering, FEI, São Bernardo do Campo, São Paulo, Brazil

Dolzycka, D., Herzmann, G., Sommer, W., & Wilhelm, O. (2014). Can training enhance face cognition abilities in middle-aged adults? *PLOS ONE*, *9*, e90249. doi: 10.1371/journal.pone.0090249

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, *106*, 433–445. doi: 10.1111/bjop.12103

Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *17*, 249–261. doi: 10.1162/0898929053124857

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*, 576–585. doi: 10.1016/j.neuropsychologia.2005.07.001

Duchaine, B., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, *41*, 713–720. doi: 10.1016/S0028-3932(02)00222-1

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society*, *2*, 71–73. doi: 10.3758/BF03327717

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*, 431–439. doi: 10.1068/p080431

Fast, K., Fujiwara, E., & Markowitsch, H. J. (2004). Famous Faces Test. *Göttingen: Hogrefe*.

Faundez-Zanuy, M. (2005). Data fusion in biometrics. *IEEE Aerospace and Electronic Systems Magazine*, *20*, 34–38. doi: 10.1109/MAES.2005.1396793

Fletcher, K. I., Butavicius, M. A., & Lee, M. D. (2008). Attention to internal face features in unfamiliar face matching. *British Journal of Psychology*, *99*, 379–394. doi: 10.1348/000712607X235872

Frey, C. I. (2008). *Expertise and training in identity verification* (Unpublished master's thesis). University of Zurich, Zurich.

Frowd, C., Bruce, V., McIntyre, A., & Hancock, P. (2007). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, *98*, 61–77. doi: 10.1348/000712606X104481

Fuhrer, M. & Hofer, F. (2016). *Measuring face matching and face memory accuracy of security professionals.* Manuscript submitted for publication.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, *118*, 201–210. doi: 10.1016/j.cognition.2010.11.002

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*, 211–214. doi: 10.3758/BF03329829

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*, 544–549.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. John Wiley and Sons.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*, 330–337. doi: 10.1016/S1364-6613(00)01519-9

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*, 445–464. doi: 10.1002/acp.718

Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*, 339–360. doi: 10.1037/1076-898X.5.4.339

Hengfoss, C., Mull, G., Püschel, K., & Jopp van Well, E. (2015). Herausforderungen bei der Gesichtserkennung [Challenges in face recognition]. *Kriminalistik*, *12*, 699 – 703.

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods*, *40*(3), 840–857. doi: 10.3758/BRM.40.3.840

Hildebrandt, A., Wilhelm, O., Schmiedek, F., Herzmann, G., & Sommer, W. (2011). On the specificity of face cognition compared with general cognitive functioning across adult age. *Psychology and Aging*, *26*, 701. doi: 10.1037/a0023056

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 986 - 1004. doi: 10.1037/0096-1523.22.4.986

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica*, *142*, 211–219. doi: 10.1016/j.actpsy.2012.11.013

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition*, *129*, 586–597. doi:10.1016/j.cognition.2013.08.012

Hofer, F., & Wetter, O. E. (2012). Operational and human factors issues of new airport security technology—two case studies. *Journal of Transportation Security*, *5*, 277–291. doi: 10.1007/s12198-012-0096-5

Hoogstrate, A. J., Van Den Heuvel, H., & Huyben, E. (2001). Ear identification based on surveillance camera images. *Science & Justice*, *41*, 167–172. doi: 10.1016/S1355-0306(01)71885-0

Hurley, D. J., Nixon, M. S., & Carter, J. N. (2005). Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, *98*, 491–512. doi: 10.1016/j.cviu.2004.11.001

Iannarelli, A. V. (1989). *Ear identification*. Paramont Publishing Company.

Imhofer, R. (1906). Die Bedeutung der Ohrmuschel für die Feststellung der Identität. [The importance of the external ear for identification]. *Archiv Für Die Kriminologie*, *26*, 3.

International Civil Aviation Organization (2015). *Machine readable documents* (7th Edition, Doc 9303). Retrieved from: http://www.icao.int/security/mrtd

Interpol. (n.d.). *Wanted Persons*. Retrieved from http://www.interpol.int/notice/search/wanted

Introna, L., & Nissenbaum, H. (2010). *Facial Recognition Technology A Survey of Policy and Implementation Issues*. Retrieved from http://eprints.lancs.ac.uk/49012/

iView X (Version 2.8) [Computer software]. Teltow, Germany: SensoMotoric Instrumens

Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*, 4–20. doi: 10.1109/TCSVT.2003.818349

Jain, V., & Mukherjee, A (2002). The indian face database. Retrieved from http://vis-www.cs.umass.edu/vidit/IndianFaceDatabase.

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2-9. doi: 10.7771/1932-6246.1167

Jaslow, R. (2013, September 27). London police using 200 super-recognizers: What makes them super? *CBS News.* Retrieved from http://www.cbsnews.com/news/london-police-using-200-super-recognizers-what-makes-them-super/

JASP Team. (2016). JASP (Version 0.7.5.5.).

Jeffreys, H. (2000). *Theory of Probability* (Third.). Oxford Oxfordshire : New York: Oxford University Press.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*, 577–596. doi: 10.1080/09658210902976969

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*, 211–222. doi: 10.1002/(SICI)1099-0720(199706)11:3<211::AID-ACP430>3.0.CO;2-O

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics*, *140A*, 1617–1622. doi: 10.1002/ajmg.a.31343

Kleinberg, K. F., Vanezis, P., & Burton, A. M. (2007). Failure of anthropometry as a facial identification technique using high-quality photographs. *Journal of Forensic Sciences*, *52*, 779–783. doi: 10.1111/j.1556-4029.2007.00458.x

Klontz, J. C., & Jain, A. K. (2013). A case study of automated face recognition: The Boston marathon bombings suspects. *Computer*, *46*, 91–94. doi: 10.1109/MC.2013.377

Koller, C. I., Wetter, O. E., & Hofer, F. (2015a). What is suspicious when trying to be inconspicuous? Criminal intentions inferred from nonverbal behavioral cues. *Perception*, *44*, 679–708. doi: 10.1177/0301006615594271

Koller, C. I., Wetter, O. E., & Hofer, F. (2015b). Who's the thief? The influence of knowledge and experience on early detection of criminal intentions. *Applied Cognitive Psychology*, *30*, 178–187. doi: 10.1002/acp.3175

Kumar, A., & Wu, C. (2012). Automated human identification using ear imaging. *Pattern Recognition*, *45*, 956–968. doi: 10.1016/j.patcog.2011.06.005

Labati, R. D., Genovese, A., Muñoz, E., Piuri, V., Scotti, F., & Sforza, G. (2015). Advanced design of Automated Border Control gates: Biometric system techniques and research trends. In *2015 IEEE International Symposium on Systems Engineering (ISSE)* (pp. 412–419). doi: 10.1109/SysEng.2015.7302791

Lee, M. D., Vast, R. L., & Butavicius, M. A. (2006). *Face matching under time pressure and task demands*. In R. Sun (Ed.), Proceedings of the 28th Annual conference of the Cognitive Science Society (pp. 1675 – 1680). Mahwah, NJ: Erlbaum.

Lindsay, R. C. L., Mansour, J. K., Bertrand, M. I., Kalmet, N., & Melsom, E. I. (2011). Face recognition in eyewitness memory. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 307 – 328). New York: Oxford University Press.

Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition*, *83*, 31–48. doi: 10.1016/S0010-0277(01)00164-0

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics*, *14*, 285–292. doi: 10.3758/BF03212392

Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2012). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior*, *36*, 513–526. doi: 10.1037/h0093937

McKone, E., & Robbins, R. (2011). Are faces special? In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook or Face Perception* (pp. 149 – 176). New York: Oxford University Press.

Megreya, A. M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception*, *38*, 1831–1848. doi: 10.1068/p6385

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, *25*, 30–37. doi: 10.1080/20445911.2012.739153

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*, 364 - 372. doi: 10.1037/a0013464

Megreya, A. M., Memon, A., & Havard, C. (2012). The Headscarf Effect: Direct Evidence from the Eyewitness Identification Paradigm. *Applied Cognitive Psychology*, *26*, 308–315. doi: 10.1002/acp.1826

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology*, *27*, 700–706. doi: 10.1002/acp.2965

Megreya, A. M., White, D., & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *The Quarterly Journal of Experimental Psychology*, *64*, 1473–1483. doi: 10.1080/17470218.2011.575228

Meinhardt-Injac, B., Meinhardt, G., & Schwaninger, A. (2009). Does matching of internal and external facial features depend on orientation and viewpoint? *Acta Psychologica*, *132*, 267–278. doi:10.1016/j.actpsy.2009.07.011

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3–35. doi: 10.1037/1076-8971.7.1.3

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*, 630–633. doi: 10.3758/BF03206543

Mythen, G., & Walklate, S. (2016). Counterterrorism and the reconstruction of (in)security: Divisions, dualisms, duplicities. *British Journal of Criminology*, azw030. doi: 10.1093/bjc/azw030

Norell, K., Läthén, K. B., Bergström, P., Rice, A., Natu, V., & O'Toole, A. J. (2015). The effect of image quality and forensic expertise in facial image comparisons. *Journal of Forensic Sciences*, *60*, 331–340. doi: 10.1111/1556-4029.12660

Oostveen, A.-M. (2014). Non-use of Automated Border Control Systems: Identifying reasons and solutions. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI* (pp. 228–233). doi: 10.14236/ewic/hci2014.28

Oostveen, A.-M., Kaufmann, M., Krempel, E., & Grasemann, G. (2014). *Automated Border Control: A Comparative Usability Study at Two European Airports* (SSRN Scholarly Paper No. ID 2432461). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2432461

Osborne, C. D., & Stevenage, S. V. (2008). Internal feature saliency as a marker of familiarity and configural processing. *Visual Cognition*, *16*, 23–43. doi: 10.1080/13506280701238073

O'Toole, A. J. (2011). Cognitive and computational approaches to face recognition. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 15 – 30). New York: Oxford University Press.

O'Toole, A. J., Phillips, P. J., Jiang, F., Ayyad, J., Penard, N., & Abdi, H. (2007). Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*, 1642–1646. doi: 10.1109/TPAMI.2007.1107

Panis, G., & Lanitis, A. (2014). An overview of research activities in facial age estimation using the FG-NET Aging Database. In L. Agapito, M. M. Bronstein, & C. Rother (Eds.), *Computer Vision - ECCV 2014 Workshops* (pp. 737–750). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-16181-5_56

Phillips, P. J., Hill, M. Q., Swindle, J. A., & O'Toole, A. J. (2015). Human and algorithm performance on the PaSC face Recognition Challenge. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)* (pp. 1–8). doi: 10.1109/BTAS.2015.7358765

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*, 1090–1104. doi: 10.1109/34.879790

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, *16*, 295–306. doi: 10.1016/S0262-8856(97)00070-X

Purkait, R. (2016). External ear: An analysis of its uniqueness. *Egyptian Journal of Forensic Sciences*. doi:10.1016/j.ejfs.2016.03.002

Purkait, R., & Singh, P. (2008). A test of individuality of human external ear pattern: Its application in the field of personal identification. *Forensic Science International*, *178*, 112–118. doi: 10.1016/j.forsciint.2008.02.009

Raghavendra, R., Raja, K. B., & Busch, C. (2015). Presentation attack fetection for face recognition using light field camera. *IEEE Transactions on Image Processing*, *24*, 1060–1075. doi: 10.1109/TIP.2015.2395951

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Revelle, W. (2015). Psych: Procedures for personality and psychological research (Version 1.5.8). Retrieved from http://CRAN.R-project.org/package=psych

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*, 146–174. doi: 10.1037/a0025750

Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, *20*, 143–169. doi: 10.1080/13506285.2012.654624

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face Recognition by Metropolitan Police Super-Recognisers. *PLOS ONE*, *11*, e0150036. doi: 10.1371/journal.pone.0150036

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of statistical software, 48,* 1 – 36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience (Elite Edition)*, *6*, 258–307.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*, 252–257. doi: 10.3758/PBR.16.2.252

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*, 23–74.

Security Newsdesk (2013). BSIA attempts to clarify question of how many CCTV cameras there are in the UK. Retrieved from http://www.securitynewsdesk.com/

Somanath, G., Rohith, M., & Kambhamettu, C. (2011). VADANA: A dense dataset for facial image analysis. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 2175–2182). doi: 10.1109/ICCVW.2011.6130517

Schwaninger, A. (2005). Increasing efficiency in airport security screening. *WIT Transactions on the Built Environment*, *82*. doi: 10.2495/SAFE050401

Schwaninger, A., Hofer, F., & Wetter, O. E. (2007). Adaptive computer-based training increases on the job performance of X-Ray screeners. In *2007 41st Annual IEEE International Carnahan Conference on Security Technology* (pp. 117–124). doi: 10.1109/CCST.2007.4373478

Schwaninger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. In H. H. Bülthoff, C. Wallraven, S.-W. Lee, & T. A. Poggio (Eds.), *Biologically Motivated Computer Vision* (pp. 643–650). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/3-540-36181-2_64

Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition*, *27*, 538–552. doi: 10.3758/BF03211547

Spreeuwers, L. J., Hendrikse, A. J., & Gerritsen, K. J. (2012). Evaluation of automatic face recognition for automatic border control on actual data recorded of travellers at Schiphol Airport. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the* (pp. 1–6).

Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: an event-related potential study: *NeuroReport*, *19*, 583–587. doi: 10.1097/WNR.0b013e3282f97b4d

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*, 225–245. doi: 10.1080/14640749308401045

Tanaka, J. W., & Gordon, I. (2011). Features, configuration, and holistic processing. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception*. New York: Oxford University Press.

Toseeb, U., Bryant, E. J., & Keeble, D. R. T. (2014). The muslim headscarf and face perception: They all look the same, don't they? *PLOS ONE*, *9*, e84754. doi: 10.1371/journal.pone.0084754

Towler, A., White, D., & Kemp, R. I. (2014). Evaluating training methods for facial image comparison: The face shape strategy does not work. *Perception*, *43*, 214–218. doi: 10.1068/p7676

University of Stirling Psychology Department (n.d.) *Psychological image collection at Stirling (PICS)*. Retrieved from http://pics.psych.stir.ac.uk/

U. S. Federal Bureau of Investigation (n.d.). *Most wanted. Be part of the solution*. Retrieved from https://www.fbi.gov/wanted

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*, 161–204. doi: 10.1080/14640749108400966

Van Belle, G., Ramon, M., Lefèvre, P., Rossion, B., & others. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. *Frontiers in Psychology*, *1*, 1–8. doi: 10.3389/fpsyg.2010.00020

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review*, *14*, 779–804. doi: 10.3758/BF03194105

Wang, Y., Thomas, J., Weissgerber, S. C., Kazemini, S., Ul-Haq, I., & Quadflieg, S. (2015). The headscarf effect revisited: Further evidence for a culture-based internal face processing advantage. *Perception*, *44*, 328–336. doi: 10.1068/p7940

Warrington, E. K. (1984). *Recognition memory test.* Windsor (UK): NFER-Nelson.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 00–00. doi: 10.1080/00031305.2016.1154108

Wetter, O. E. (2012). *From short-term goals to organizational objectives: Managing security control using goal setting* (Doctoral dissertation). Retrieved from http://www.dissertationen.uzh.ch/

Wetter, O. E., & Fuhrer, M. (2013). A holistic approach for evaluating liquid explosive detection systems. *Journal of Transportation Security*, *6*, 377–388. doi: 10.1007/s12198-013-0122-2

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error Rates in Users of Automatic Face Recognition Software. *PLOS ONE*, *10*, e0139827. doi: 10.1371/journal.pone.0139827

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*, 100–106. doi: 10.3758/s13423-013-0475-3

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS One*, *9*, e103510. doi: 10.1371/journal.pone.0103510

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B, 282*(1814), 20151292. doi: 10.1098/rspb.2015.1292

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces: One element of social cognition. *Journal of Personality and Social Psychology*, *99*, 530–548. doi: 10.1037/a0019972

Wilkinson, C., & Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science & Justice*, *49*, 191–196. doi: 10.1016/j.scijus.2008.10.011

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: The example of face recognition. *Cognitive Neuropsychology*, *29*, 360–392. doi: 10.1080/02643294.2012.753433

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*, 5238–5241. doi: 10.1073/pnas.0913053107

Woodhead, M. M., & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & Cognition*, *9*, 368–370. doi: 10.3758/BF03197561

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognize faces. *Ergonomics*, *22*, 333–343. doi: 10.1080/00140137908924617

Yan, P., & Bowyer, K. (2005). Empirical evaluation of advanced ear biometrics. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on* (pp. 41–41). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1565339

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, *31*, 995–1003. doi: 10.1068/p3376

Zhou, X., & Li, B. (2015). Retrieving unfamiliar faces: Towards understanding human performance. In *Proceedings of the 23rd ACM International Conference on Multimedia* (pp. 1287–1290). New York, NY, USA: ACM. doi: 10.1145/2733373.2806381

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., … Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*, *20*, 137–142. doi: 10.1016/j.cub.2009.11.067

Acknowledgement

Without some people, this work would not have been possible.

First, I would like to thank Franziska Hofer for her supervision. Thank you for your advice and support, for the time you took for our exchanges, as well as your understanding and openness in all respects. I benefited so much from your experience and knowledge!

I sincerely thank Klaus Oberauer, for his advice whenever I needed it, for all the valuable inputs and for the freedom he gave me as a PhD student. I thank Janek Lobmaier for being willing to be my second supervisor, it was always agreeable and interesting when we met.

Thank you Corinne Koller, for the contribution of your ideas, for the understanding you always had because we were on the same boat and for your advice and your funny comments. I thank also my PhD-student colleagues Carla DeSimoni, Marcel Niklaus and Mirko Thalmann who shared one office with me. I benefited a lot from our discussions! Further, I would like to thank all of my research interns, tutors and students, especially Amanda Planzer, Christina Tschan and Florian Schuppisser for your work, your ideas, your motivation and your assistance with several tasks! Thanks also to the whole cognition team for your helpful comments and questions in the FoKo and the uncomplicated cooperation.

I thank the Kantonspolizei Zürich, the Federal Office of Civil Aviation and the University of Zurich for the opportunity to do a dissertation in a practical field.
Special thanks go to Franz Bättig, who had the lead of this project within the police. Thank you for your inputs from the applied side, your openness, as well as the

opportunity to teach within the police, that was an invaluable experience.
I would like to thank Andi, Martin and Roli for the exchange, the cooperation in ASPECT workshops and that you were willing to give our students an insight into the world outside university.
It was great to visit the Metropolitan Police in London with you, Martin and Mike, thank you for these interesting and funny days. Thanks also to the rest of the ASPECT-team, it was always a pleasure to meet you.

Furthermore, I would like to thank the people supporting me beyond work.

I am very grateful to my friends who share thoughts and time with me and accompany me in good and difficult times. I don't know where I would be without you and I don't have words to tell how much you mean to me. Special thanks go to my very closest friends Rea Lauper, Mara Heer, Nina Wolf, Romedi Zegg, Severin Stadtmann and Dominik Gsponer.
Thanks go also to my family for their support. Martin Fuhrer, you are the best brother in the world and I don't know how many times you saved me with your sense of humour. I thank Beat Javet, my father, for being how you are and for believing in me and my abilities through my whole life. Thank you Erica Ramsauer and Denise Javet, Simon and Colin Wright for being family without being related by blood!

Curriculum Vitae of Mirjam Fuhrer

Contact	mirjam.fuhrer@uzh.ch
Birth date	18.01.1983
Place of origin	Langnau im Emmental (BE)

Academic Studies

2013 – 2016	PhD Program Psychology, University of Zurich Supervisors: Prof. Dr. Klaus Oberauer and Prof. Dr. Janek Lobmaier Doctoral Thesis: Assessment and improvement of unfamiliar face processing and identity verification in security and forensics
2010 - 2013	Master of Science, University of Zurich, Psychology (major), Sociology (minor) Master Thesis: Der Einfluss von Vorstellungen über Täuschung auf die Wahrnehmung von nonverbalem Verhalten [The influence of beliefs about deception on the perception of nonverbal behavior]
2006 – 2010	Bachelor of Science, University of Zurich, Psychology (major), Sociology and Biology (minors)

Professional Experience

2013 – 2016	University of Zurich, Department of Psychology, Cognitive Psychology Unit Assistant and PhD-Student
2013 – 2016	Kantonspolizei Zürich, Instructor ASPECT Workshops
2013	Kantonspolizei Zürich, Airport Division, Research & Development Research associate
2008 – 2013	Psychiatrische Universitätsklinik Zürich Nursing assistant
2012	Kantonspolizei Zürich, Airport Division, Research & Development Research Intern

Publications

Fuhrer, M. & Hofer, F. (2016). Attention to the ears improves simultaneous face matching accuracy: A simple method to enhance performance in security and forensic settings. *Manuscript submitted for publication.*

Fuhrer, M. & Hofer, F. (2016). Measuring face matching and face memory performance of security professionals. *Manuscript submitted for publication.*

Koller, C. I. & Fuhrer, M. (2016). Was tun, wenn die Pinocchio-Nase nicht funktioniert? Vom Versuch der Lügenerkennung anhand nonverbalen Verhaltens hin zu Befragungstechniken. *Manuscript submitted for publication.*

Wetter, O. E. & Fuhrer, M. (2013). A holistic approach for evaluating liquid explosive detection systems. *Journal of Transportation Security, 6,* 377 - 388.

Universität Zürich[UZH]

Philosophische Fakultät
Studiendekanat

Universität Zürich
Studiendekanat
Rämistrasse 69
CH-8001 Zürich
www.phil.uzh.ch

Studienabschlüsse
abschluss@phil.uzh.ch

Erklärung

Hiermit erkläre ich, dass
- die Dissertation von mir selbst ohne unerlaubte Beihilfe verfasst worden ist und
- diese Dissertation noch an keiner anderen Fakultät eingereicht wurde.

Ort und Datum Unterschrift

Zürich, 28.7.2016